# **PCT**

# WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



### INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>5</sup>: C12N 15/51, C12Q 1/68 C12N 15/40, C12Q 1/70 A61K 39/29, C07K 13/00 G01N 33/576

(11) International Publication Number:

WO 92/19743

**A2** 

(43) International Publication Date:

12 November 1992 (12.11.92)

(21) International Application Number:

PCT/US92/04036

(22) International Filing Date:

8 May 1992 (08.05.92)

(30) Priority data:

697,326

8 May 1991 (08.05.91)

US

(71) Applicant: CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors: CHA, Tai-An; 964 Springview Circle, San Ramon, CA 94583 (US). BEALL, Eileen; 1150 Lincoln Avenue, #5, Walnut Creek, CA 94596 (US). IRVINE, Bruce; 3401 El Monte Drive, Concord, CA 94519 (US). KOLBERG, Janice; 131 Scots Valley, Hercules, CA 94547 (US). URDEA, Michael, S.; 100 Bunce Meadow Road, Alamo, CA 94501 (US).

(74) Agent: JANIUK, Anthony, J.; Wolf, Greenfield & Sacks, 600 Atlantic Avenue, Boston, MA 02210 (US).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, RU, SD, SE (European patent), SN (OAPI patent), TD (OAPI patent), TG (OAPI patent).

#### **Published**

Without international search report and to be republished upon receipt of that report.

(54) Title: HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

#### (57) Abstract

The present application features nucleic acid, peptide and antibody compositions relating to genotypes of hepatitis C virus and methods of using such compositions for diagnostic and therapeutic purposes.

# FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT AU BB BE BF BC BJ BR CA CF CG CH CM CS DE DK	Australia Barbailos Belgium Burkina Faso Bulgaria Benin Brazil Canada Central African Republic Congo Switzerland C'âte d'Ivoire Cameroon Czechuslovakia Germany Denmark	ES FI FR GA GB GN GR HU IT JP KP KR LI LK LU MC	Spain Finland France Gabon United Kingdom Guinea Greece Hungary Italy Japan Democratic People's Republic of Korea Republic of Korea Liechtenstein Sri Lanka Lusembourg Monaco	MG. MI. MN MR MW NL NO PL RO RU SD SE SN SU TD TG US	Madagascar Mali Mongolia Mauritania Malawi Netherlands Norway Poland Romania Russian Federation Sudan Sweden Senegal Soviet Union Chad Togo United States of America	
---	---	--	---	--	--	--

1

# HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

This application is a continuation-in-part of U.S. Serial No. 07/697,326 entitled "Polynucleotide Probes Useful for Screening for Hepatitis C Virus, filed May 8, 1991.

### Technical Field

The invention relates to compositions and methods for the detection and treatment of hepatitis C virus, (HCV) infection, formerly referred to as blood-borne non-A, non-B hepatitis virus (NANBV) infection. More specifically, embodiments of the present invention feature compositions and methods for the detection of HCV, and for the development of vaccines for the prophylactic treatment of infections of HCV, and development of antibody products for conveying passive immunity to HCV.

20

25

# Background of the Invention

The prototype isolate of HCV was characterized in U.S. Patent Application Serial No. 122,714 (See also EPO Publication No. 318,216). As used herein, the term "HCV" includes new isolates of the same viral species. The term "HCV-1" referred to in U.S. Patent Application Serial No. 122,714.

- 2 -

HCV is a transmissible disease distinguishable from other forms of viral-associated liver diseases, including that caused by the known hepatitis viruses, i.e., hepatitis A virus (HAV), hepatitis B virus (HBV), and delta hepatitis virus (HDV), as well as the hepatitis induced by cytomegalovirus (CMV) or Epstein-Barr virus (EBV). HCV was first identified in individuals who had received blood transfusions.

5

10

15

20

25

The demand for sensitive, specific methods for screening and identifying carriers of HCV and HCV contaminated blood or blood products is significant. Post-transfusion hepatitis (PTH) occurs in approximately 10% of transfused patients, and HCV accounts for up to 90% of these cases. The disease frequently progresses to chronic liver damage (25-55%).

Patient care as well as the prevention of transmission of HCV by blood and blood products or by close personal contact require reliable screening, diagnostic and prognostic tools to detect nucleic acids, antigens and antibodies related to HCV.

Information in this application suggests the HCV has several genotypes. That is, the genetic information of the HCV virus may not be totally identical for all HCV, but encompasses groups with differing genetic information.

Genetic information is stored in thread-like molecules of DNA and RNA. DNA consists of covalently

linked chains of deoxyribonucleotides and RNA consists of covalently linked chains of ribonucleotides. Each nucleotide is characterized by one of four bases: adenine (A), quanine (G), thymine (T), and cytosine (C). The bases are complementary in the sense that, 5 due to the orientation of functional groups, certain base pairs attract and bond to each other through hydrogen bonding and  $\pi$ -stacking interactions. Adenine in one strand of DNA pairs with thymine in an opposing complementary strand. Guanine in one strand 10 of DNA pairs with cytosine in an opposing complementary strand. In RNA, the thymine base is replaced by uracil (U) which pairs with adenine in an opposing complementary strand. The genetic code of living organism is carried in the sequence of base pairs. 15 Living cells interpret, transcribe and translate the information of nucleic acid to make proteins and peptides.

The HCV genome is comprised of a single positive
strand of RNA. The HCV genome possesses a continuous,
translational open reading frame (ORF) that encodes a
polyprotein of about 3,000 amino acids. In the ORF,
the structural protein(s) appear to be encoded in
approximately the first quarter of the N-terminus
region, with the majority of the polyprotein
responsible for non-structural proteins.

- 4 -

The HCV polyprotein comprises, from the amino terminus to the carboxy terminus, the nucleocapsid protein (C), the envelope protein (E), and the non-structural proteins (NS) 1, 2 (b), 3, 4 (b), and 5.

HCV of differing genotypes may encode for proteins which present an altered response to host immune systems. HCV of differing genotypes may be difficult to detect by immuno diagnostic techniques and nucleic acid probe techniques which are not specifically directed to such genotype.

10

15

20

Definitions for selected terms used in the application are set forth below to facilitate an understanding of the invention. The term "corresponding" means homologous to or complementary to a particular sequence of nucleic acid. As between nucleic acids and peptides, corresponding refers to amino acids of a peptide in an order derived from the sequence of a nucleic acid or its complement.

The term "non-naturally occurring nucleic acid" refers to a portion of genomic nucleic acid, cDNA, semisynthetic nucleic acid, or synthetic origin nucleic acid which, by virtue of its origin or manipulation:

(1) is not associated with all of a nucleic acid with which it is associated in nature, (2) is linked to a nucleic acid or other chemical agent other than that to

- 5 -

which it is linked in nature, or (3) does not occur in nature.

Similarly the term, "a non-naturally occurring peptide" refers to a portion of a large naturally occurring peptide or protein, or semi-synthetic or synthetic peptide, which by virtue of its origin or manipulation (1) is not associated with all of a peptide with which it is associated in nature, (2) is linked to peptides, functional groups or chemical agents other than that to which it is linked in nature, or (3) does not occur in nature.

10

15

20

25

The term "primer" refers to a nucleic acid which is capable of initiating the synthesis of a larger nucleic acid when placed under appropriate conditions. The primer will be completely or substantially complementary to a region of the nucleic acid to be copied. Thus, under conditions conducive to hybridization, the primer will anneal to a complementary region of a larger nucleic acid. Upon addition of suitable reactants, the primer is extended by the polymerizing agent to form a copy of the larger nucleic acid.

The term "binding pair" refers to any pair of molecules which exhibit mutual affinity or binding capacity. For the purposes of the present application, the term "ligand" will refer to one molecule of the binding pair, and the term "antiligand" or "receptor"

- 6 -

or "target" will refer to the opposite molecule of the binding pair. For example, with respect to nucleic acids, a binding pair may comprise two complementary nucleic acids. One of the nucleic acids may be designated the ligand and the other strand is designated the antiligand receptor or target. The designation of ligand or antiligand is a matter of arbitrary convenience. Other binding pairs comprise, by way of example, antigens and antibodies, drugs and drug receptor sites and enzymes and enzyme substrates, to name a few.

The term "label" refers to a molecular moiety capable of detection including, by way of example, without limitation, radioactive isotopes, enzymes, luminescent agents, precipitating agents, and dyes.

10

15

20

The term "support" includes conventional supports such as filters and membranes as well as retrievable supports which can be substantially dispersed within a medium and removed or separated from the medium by immobilization, filtering, partitioning, or the like. The term "support means" refers to supports capable of being associated to nucleic acids, peptides or antibodies by binding partners, or covalent or noncovalent linkages.

A number of HCV strains and isolates have been identified. When compared with the sequence of the original isolate derived from the USA ("HCV-1"; see

- 7 -

O.-L. Choo et al. (1989) Science 244:359-362, Q.-L. Choo et al. (1990) Brit. Med. Bull. 46:423-441, Q.-L. Choo et al., Proc. Natl. Acad. Sci. 88:2451-2455 (1991), and E.P.O. Patent Publication No. 318,216, cited supra), it was found that a Japanese isolate ("HCV J1") differed significantly in both nucleotide and polypeptide sequence within the NS3 and NS4 regions. This conclusion was later extended to the NS5 and envelope (E1/S and E2/NS1) regions (see K. Takeuchi et al., J. Gen. Virol. (1990) 71:3027-3033, Y. Kubo, 10 Nucl. Acids. Res. (1989) 17:10367-10372, and K. Takeuchi et al., Gene (1990) 91:287-291). The former group of isolates, originally identified in the United States, is termed "Genotype I" throughout the present disclosure, while the latter group of isolates, 15 initially identified in Japan, is termed "Genotype II" herein.

### Brief Description of the Invention

The present invention features compositions of matter comprising nucleic acids and peptides corresponding to the HCV viral genome which define different genotypes. The present invention also features methods of using the compositions corresponding to sequences of the HCV viral genome which define different genotypes described herein.

- 8 -

# A. Nucleic acid compositions

5

10

15

20

25

The nucleic acid of the present invention, corresponding to the HCV viral genome which define different genotypes, have utility as probes in nucleic acid hybridization assays, as primers for reactions involving the synthesis of nucleic acid, as binding partners for separating HCV viral nucleic acid from other constituents which may be present, and as anti-sense nucleic acid for preventing the transcription or translation of viral nucleic acid.

One embodiment of the present invention features a composition comprising a non-naturally occurring nucleic acid having a nucleic acid sequence of at least eight nucleotides corresponding to a non-HCV-1 nucleotide sequence of the hepatitis C viral genome. Preferably, the nucleotide sequence is selected from a sequence present in at least one region consisting of the NS5 region, envelope 1 region, 5'UT region, and the core region.

Preferably, with respect to sequences which correspond to the NS5 region, the sequence is selected from a sequence within a sequence numbered 2-22. The sequence numbered 1 corresponds to HCV-1. Sequences numbered 1-22 are defined in the Sequence Listing of the application.

Preferably, with respect to sequences corresponding to the envelope 1 region, the sequence is

- 9 -

selected from a sequence within sequences numbered 24-32. Sequence No. 23 corresponds to HCV-1. Sequences numbered 23-32 are set forth in the Sequence Listing of the application.

5

10

15

20

25

Preferably, with respect to the sequences which correspond to the 5'UT regions, the sequence is selected from a sequence within sequences numbered 34-51. Sequence No. 33 corresponds to HCV-1. Sequence No. 33-51 are set forth in the Sequence Listing of this application.

Preferably, with respect to the sequences which correspond to the core region, the sequence is selected from a sequence within the sequences numbered 53-66. Sequence No. 52 corresponds to HCV-1. Sequences 52-66 are set forth in the Sequence Listing of this application.

The compositions of the present invention form hybridization products with nucleic acid corresponding to different genotypes of HCV.

HCV has at least five genotypes, which will be referred to in this application by the designations GI-GV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV,

10

25

is exemplified by sequences numbered 20-22, and 29-31 and 48-49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

One embodiment of the present invention features compositions comprising a nucleic acid having a sequence corresponding to one or more sequences which exemplify a genotype of HCV.

B. Method of forming a Hybridization Product

Embodiments of the present invention also feature a method of forming a hybridization product with nucleic acid having a sequence corresponding to HCV nucleic acid. One method comprises the steps of placing a non-naturally occurring nucleic acid having a non-HCV-1 sequence corresponding to HCV nucleic acid 15 under conditions in which hybridization may occur. The non-naturally occurring nucleic acid is capable of forming a hybridization product with HCV nucleic acid, under hybridization conditions. The method further comprises the step of imposing hybridization conditions to form a hybridization product in the presence of nucleic acid corresponding to a region of the HCV genome.

The formation of a hybridization product has utility for detecting the presence of one or more genotypes of HCV. Preferably, the non-naturally occurring nucleic acid forms a hybridization product PCT/US92/04036

WO 92/19743

10

15

20

25

with nucleic acid of HCV in one or more regions comprising the NS5 region, envelope 1 region, 5'UT region and the core region. To detect the hybridization product, it is useful to associate the non-naturally occurring nucleic acid with a label. The formation of the hybridization product is detected by separating the hybridization product from labeled non-naturally occurring nucleic acid, which has not formed a hybridization product.

The formation of a hybridization product has utility as a means of separating one or more genotypes of HCV nucleic acid from other constituents potentially present. For such applications, it is useful to associate the non-naturally occurring nucleic acid with a support for separating the resultant hybridization product from the the other constituents.

Nucleic acid "sandwich assays" employ one nucleic acid associated with a label and a second nucleic acid associated with a support. An embodiment of the present invention features a sandwich assay comprising two nucleic acids, both have sequences which correspond to HCV nucleic acids; however, at least one non-naturally occurring nucleic acid has a sequence corresponding to non-HCV-1 HCV nucleic acid. At least one nucleic acid is capable of associating with a label, and the other is capable of associating with a support. The support associated non-naturally

25

occurring nucleic acid is used to separate the hybridization products which include an HCV nucleic acid and the non-naturally occurring nucleic acid having a non-HCV-1 sequence.

One embodiment of the present invention features a 5 method of detecting one or more genotypes of HCV. The method comprises the steps of placing a non-naturally occurring nucleic acid under conditions which hybridization may occur. The non-naturally occurring nucleic acid is capable of forming a hybridization 10 product with nucleic acid from one or more genotypes of The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third 15 genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified sequences numbered 20-22 and 29-31. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51. 20

The hybridization product of HCV nucleic acid with a non-naturally occurring nucleic acid having non-HCV-1 sequence corresponding to sequences within the HCV genome has utility for priming a reaction for the synthesis of nucleic acid.

The hybridization product of HCV nucleic acid with a non-naturally occurring nucleic acid having a

- 13 -

sequence corresponding to a particular genotype of HCV has utility for priming a reaction for the synthesis of nucleic acid of such genotype. In one embodiment, the synthesized nucleic acid is indicative of the presence of one or more genotypes of HCV.

The synthesis of nucleic acid may also facilitate cloning of the nucleic acid into expression vectors which synthesize viral proteins.

Embodiments of the present methods have utility as anti-sense agents for preventing the transcription or translation of viral nucleic acid. The formation of a hybridization product of a non-naturally occurring nucleic acid having sequences which correspond to a particular genotype of HCV genomic sequencing with HCV nucleic acid may block translation or transcription of such genotype. Therapeutic agents can be engineered to include all five genotypes for inclusivity.

### C. Peptide and antibody composition

10

15

A further embodiment of the present invention

features a composition of matter comprising a
non-naturally occurring peptide of three or more amino
acids corresponding to a nucleic acid having a
non-HCV-1 sequence. Preferably, the non-HCV-1 sequence
corresponds with a sequence within one or more regions
consisting of the NS5 region, the envelope 1 region,
the 5'UT region, and the core region.

Listing.

10

15

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence of the NS5 region, the sequence is within sequences numbered 2-22. The sequence numbered 1 corresponds to HCV-1. Sequences numbered 1-22 are set forth in the Sequence

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence of the envelope 1 region, the sequence is within sequences numbered 24-32. The sequence numbered 23 corresponds to HCV-1. Sequences numbered 23-32 are set forth in the Sequence Listing.

Preferably, with respect to peptides corresponding to a nucleic acid having a non-HCV-1 sequence directed to the core region, the sequence is within sequences numbered 53-66. Sequence numbered 52 corresponds to HCV-1. Sequences numbered 52-66 are set forth in the Sequence Listing.

The further embodiment of the present invention

features peptide compositions corresponding to nucleic acid sequences of a genotype of HCV. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified

- 15 -

sequences numbered 20-22, 29-31, 48 and 49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

5

10

15

20

25

The non-naturally occurring peptides of the present invention are useful as a component of a vaccine. The sequence information of the present invention permits the design of vaccines which are inclusive for all or some of the different genotypes of HCV. Directing a vaccine to a particular genotype allows prophylactic treatment to be tailored to maximize the protection to those agents likely to be encountered. Directing a vaccine to more than one genotype allows the vaccine to be more inclusive.

The peptide compositions are also useful for the development of specific antibodies to the HCV proteins. One embodiment of the present invention features as a composition of matter, an antibody to peptides corresponding to a non-HCV-1 sequence of the HCV genome. Preferably, the non-HCV-1 sequence is selected from the sequence within a region consisting of the NS5 region, the envelope 1 region, and the core region. There are no peptides associated with the untranslated 5'UT region.

Preferably, with respect to antibodies directed to peptides of the NS5 region, the peptide corresponds to a sequence within sequences numbered 2-22. Preferably, with respect to antibodies directed to a peptide

- 16 -

corresponding to the envelope 1 region, the peptide corresponds to a sequence within sequences numbered 24-32. Preferably, with respect to the antibodies directed to peptides corresponding to the core region, the peptide corresponds to a sequence within sequences numbered 53-66.

5

10

15

20

25

Antibodies directed to peptides which reflect a particular genotype have utility for the detection of such genotypes of HCV and therapeutic agents.

One embodiment of the present invention features an antibody directed to a peptide corresponding to nucleic acid having sequences of a particular genotype. The first genotype, GI, is exemplified by sequences numbered 1-6, 23-25, 33-38 and 52-57. The second genotype, GII, is exemplified by the sequences numbered 7-12, 26-28, 39-45 and 58-64. The third genotype, GIII, is exemplified by sequences numbered 13-17, 32, 46-47 and 65-66. The fourth genotype, GIV, is exemplified sequences numbered 20-22, 29-31, 48 and 49. The fifth genotype, GV, is exemplified by sequences numbered 18, 19, 50 and 51.

Individuals skilled in the art will readily recognize that the compositions of the present invention can be packaged with instructions for use in the form of a kit for performing nucleic acid hybridizations or immunochemical reactions.

PCT/US92/04036

The present invention is further described in the following figures which illustrate sequences demonstrating genotypes of HCV. The sequences are designated by numerals 1-145, which numerals and sequences are consistent with the numerals and sequences set forth in the Sequence Listing. Sequences 146 and 147 facilitate the discussion of an assay which numerals and sequences are consistent with the numerals and sequences set forth in the Sequence Listing.

10

15

20

25

5

Brief Description of the Figures and Sequence Listing
Figure 1 depicts schematically the genetic
organization of HCV;

Figure 2 sets forth nucleic acid sequences numbered 1-22 which sequences are derived from the NS5 region of the HCV viral genome;

Figure 3 sets forth nucleic acid sequences numbered 23-32 which sequences are derived from the envelope 1 region of the HCV viral genome;

Figure 4 sets forth nucleic acid sequences numbered 33-51 which sequences are derived from the 5'UT region of the HCV viral genome; and,

Figure 5 sets forth nucleic acid sequences numbered 52-66 which sequences are derived from the core region of the HCV viral genome.

The Sequence Listing sets forth the sequences of sequences numbered 1-147.

# Detailed Description of the Invention

5

10

15

The present invention will be described in detail as as nucleic acid having sequences corresponding to the HCV genome and related peptides and binding partners, for diagnostic and therapeutic applications.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitsch & Sambrook, Molecular Cloning; A Laboratory Manual (1982); DNA Cloning, Volumes I and II (D.N Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed, 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); the series, Methods in Enzymology (Academic Press, Inc.), particularly Vol. 154 and Vol. 155 (Wu and Grossman, eds.).

The cDNA libraries are derived from nucleic acid
sequences present in the plasma of an HCV-infected
chimpanzee. The construction of one of these
libraries, the "c" library (ATCC No. 40394), is
described in PCT Pub. No. WO90/14436. The sequences of
the library relevant to the present invention are set
forth herein as sequence numbers 1, 23, 33 and 52.

Nucleic acids isolated or synthesized in accordance with features of the present invention are

- 19 -

useful, by way of example without limitation as probes, primers, anti-sense genes and for developing expression systems for the synthesis of peptides corresponding to such sequences.

The nucleic acid sequences described define genotypes of HCV with respect to four regions of the viral genome. Figure 1 depicts schematically the organization of HCV. The four regions of particular interest are the NS5 region, the envelope 1 region, the 5'UT region and the core region.

5

10

15

20

25

The sequences set forth in the present application as sequences numbered 1-22 suggest at least five genotypes in the NS5 region. Sequences numbered 1-22 are depicted in Figure 2 as well as the Sequence Listing. Each sequence numbered 1-22 is derived from nucleic acid having 340 nucleotides from the NS5 region.

The five genotypes are defined by groupings of the sequences defined by sequence numbered 1-22. For convenience, in the present application, the different genotypes will be assigned roman numerals and the letter "G".

The first genotype (GI) is exemplified by sequences within sequences numbered 1-6. A second genotype (GII) is exemplified by sequences within sequences numbered 7-12. A third genotype (GIII) is exemplified by the sequences within sequences numbered 13-17. A fourth genotype (GIV) is exemplified by

PCT/US92/04036 WO 92/19743

- 20 -

sequences within sequences numbered 20-22. A fifth genotype (GV) is exemplified by sequences within sequences numbered 18 and 19.

5

10

15

The sequences set forth in the present application as sequences numbered 23-32 suggest at least four genotypes in the envelope 1 region of HCV. Sequences numbered 23-32 are depicted in Figure 3 as well as in the Sequence Listing. Each sequence numbered 23-32 is - derived from nucleic acid having 100 nucleotides from the envelope 1 region.

A first envelope 1 genotype group (GI) is exemplified by the sequences within the sequences numbered 23-25. A second envelope 1 genotype (GII) region is exemplified by sequences within sequences numbered 26-28. A third envelope 1 genotype (GIII) is exemplified by the sequences within sequences numbered 32. A fourth envelope 1 genotype (GIV) is exemplified by the sequences within sequence numbered 29-31.

The sequences set forth in the present application as sequences numbered 33-51 suggest at least three 20 genotypes in the 5'UT region of HCV. Sequences numbered 33-51 are depicted in Figure 4 as well as in the Sequence Listing. Each sequence numbered 33-51 is derived from the nucleic acid having 252 nucleotides from the 5'UT region, although sequences 50 and 51 are 25 somewhat shorter at approximately 180 nucleotides.

- 21 -

The first 5'UT genotype (GI) is exemplified by the sequences within sequences numbered 33-38. A second 5'UT genotype (GII) is exemplified by the sequences within sequences numbered 39-45. A third 5'UT genotype (GIII) is exemplified by the sequences within sequences numbered 46-47. A fourth 5'UT genotype (GIV) is exemplified by sequences within sequences humbered 48 and 49. A fifth 5'UT genotype (GV) is exemplified by sequences within sequences numbered 50 and 51.

The sequences numbered 48-62 suggest at least three genotypes in the core region of HCV. The sequences numbered 52-66 are depicted in Figure 5 as well as in the Sequence Listing.

10

15

20

25

The first core region genotype (GI) is exemplified by the sequences within sequences numbered 52-57. The second core region genotype (GII) is exemplified by sequences within sequences numbered 58-64. The third core region genotype (GIII) is exemplified by sequences within sequences numbered 65 and 66. Sequences numbered 52-65 are comprised of 549 nucleotides. Sequence numbered 66 is comprised of 510 nucleotides.

The various genotypes described with respect to each region are consistent. That is, HCV having features of the first genotype with respect to the NS5 region will substantially conform to features of the first genotype of the envelope 1 region, the 5'UT region and the core region.

- 22 -

Nucleic acid isolated or synthesized in accordance with the sequences set forth in sequence numbers 1-66 are useful as probes, primers, capture ligands and anti-sense agents. As probes, primers, capture ligands and anti-sense agents, the nucleic acid wil normally comprise approximately eight or more nucleotides for specificity as well as the ability to form stable hybridization products.

#### 10 Probes

15

20

A nucleic acid isolated or synthesized in accordance with a sequence defining a particular genotype of a region of the HCV genome can be used as a probe to detect such genotype or used in combination with other nucleic acid probes to detect substantially all genotypes of HCV.

With the sequence information set forth in the present application, sequences of eight or more nucleotides are identified which provide the desired inclusivity and exclusivity with respect to various genotypes within HCV, and extraneous nucleic acid sequences likely to be encountered during hybridization conditions.

Individuals skilled in the art will readily recognize that the nucleic acid sequences, for use as probes, can be provided with a label to facilitate detection of a hybridization product.

### Capture Ligand

For use as a capture ligand, the nucleic acid selected in the manner described above with respect to probes, can be readily associated with supports. The manner in which nucleic acid is associated with supports is well known. Nucleic acid having sequences corresponding to a sequence within sequences numbered 1-66 have utility to separate viral nucleic acid of one genotype from the nucleic acid of HCV of a different genotype. Nucleic acid isolated or synthesized in accordance with sequences within sequences numbered 1-66, used in combinations, have utility to capture substantially all nucleic acid of all HCV genotypes.

### 15 Primers

10

20

25

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as primers for the amplification of HCV sequences. With respect to polymerase chain reaction (PCR) techniques, nucleic acid sequences of eight or more nucleotides corresponding to one or more sequences of sequences numbered 1-66 have utility in conjunction with suitable enzymes and reagents to create copies of the viral nucleic acid. A plurality of primers having different sequences corresponding to more than one genotype can be used to create copies of viral nucleic acid for such genotypes.

10

15

20

The copies can be used in diagnostic assays to detect HCV virus. The copies can also be incorporated into cloning and expression vectors to generate polypeptides corresponding to the nucleic acid synthesized by PCR, as will be described in greater detail below.

### Anti-sense

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as anti-sense genes to prevent the expression of HCV.

Nucleic acid corresponding to a genotype of HCV is loaded into a suitable carrier such as a liposome for introduction into a cell infected with HCV. A nucleic acid having eight or more nucleotides is capable of binding to viral nucleic acid or viral messenger RNA. Preferably, the anti-sense nucleic acid is comprised of 30 or more nucleotides to provide necessary stability of a hybridization product of viral nucleic acid or viral messenger RNA. Methods for loading anti-sense nucleic acid is known in the art as exemplified by U.S. Patent 4,241,046 issued December 23, 1980 to Papahadjopoulos et al.

# 25 Peptide Synthesis

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility to

- 25 -

generate peptides. The sequences exemplified by sequences numbered 1-32 and 52-66 can be cloned into suitable vectors or used to isolate nucleic acid. The isolated nucleic acid is combined with suitable DNA linkers and cloned into a suitable vector. The vector can be used to transform a suitable host organism such as <u>E. coli</u> and the peptide encoded by the sequences isolated.

Molecular cloning techniques are described in the text Molecular Cloning: A Laboratory Manual, Maniatis et al., Coldspring Harbor Laboratory (1982).

The isolated peptide has utility as an antigenic substance for the development of vaccines and antibodies directed to the particular genotype of HCV.

15

20

25

10

### Vaccines and Antibodies

The peptide materials of the present invention have utility for the development of antibodies and vaccines.

The availability of cDNA sequences, or nucleotide sequences derived therefrom (including segments and modifications of the sequence), permits the construction of expression vectors encoding antigenically active regions of the peptide encoded in either strand. The antigenically active regions may be derived from the NS5 region, envelope 1 regions, and the core region.

- 26 -

Fragments encoding the desired peptides are derived from the cDNA clones using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and vectors which are useful for the production of polypeptides which contain fusion sequences of SOD are described in European Patent Office Publication number 0196056, published October 1, 1986.

Any desired portion of the HCV cDNA containing an open reading frame, in either sense strand, can be obtained as a recombinant peptide, such as a mature or fusion protein; alternatively, a peptide encoded in the cDNA can be provided by chemical synthesis.

10

15

20

25

The DNA encoding the desired peptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, may be ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host systems are presently used in forming recombinant peptides. The peptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, differential extraction, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and

- 27 -

the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins. Such peptides can be used as diagnostics, or those which give rise to neutralizing antibodies may be formulated into vaccines. Antibodies raised against these peptides can also be used as diagnostics, or for passive immunotherapy or for isolating and identifying HCV.

5

10

15

20

An antigenic region of a peptide is generally relatively small--typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to NS5 region, envelope 1 region, and the core region of the HCV genome. The 5'UT region is not known to be translated. Accordingly, using the cDNAs of such regions, DNAs encoding short segments of HCV peptides corresponding to such regions can be expressed recombinantly either as fusion proteins, or as isolated peptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis. instances wherein the synthesized peptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the peptide may be linked to a suitable carrier.

25 A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-

pyridylthio)propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue). These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for 10 example, Immun Rev (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaprioc acid, 2-bromoacetic acid, 15 2-iodoacetic acid, 4-N-maleimido-methyl)cyclohexane-lcarboxylic acid, and the like. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2 nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the 20 rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used. 25

Any carrier may be used which does not itself induce the production of antibodies harmful to the

host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized Sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

10

Peptides comprising HCV amino acid sequences encoding at least one viral epitope derived from the NS5, envelope 1, and core region are useful immunological reagents. The 5'UT region is not known 15 to be translated. For example, peptides comprising such truncated sequences can be used as reagents in an immunoassay. These peptides also are candidate subunit antigens in compositions for antiserum production or 20 vaccines. While the truncated sequences can be produced by various known treatments of native viral protein, it is generally preferred to make synthetic or recombinant peptides comprising HCV sequence. Peptides comprising these truncated HCV sequences can be made up 25 entirely of HCV sequences (one or more epitopes, either contiguous or noncontiguous), or HCV sequences and heterologous sequences in a fusion protein. Useful

- 30 -

heterologous sequences include sequences that provide for secretion from a recombinant host, enhance the immunological reactivity of the HCV epitope(s), or facilitate the coupling of the polypeptide to an immunoassay support or a vaccine carrier. See, E.G., EPO Pub. No. 116,201; U.S. Pat. No. 4,722,840; EPO Pub. No. 259,149; U.S. Pat. No. 4,629,783.

The size of peptides comprising the truncated HCV sequences can vary widely, the minimum size being a sequence of sufficient size to provide an HCV epitope, 10 while the maximum size is not critical. For convenience, the maximum size usually is not substantially greater than that required to provide the desired HCV epitopes and function(s) of the heterologous sequence, if any. Typically, the truncated HCV amino acid sequence will range from about 5 to about 100 amino acids in length. More typically, however, the HCV sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select HCV 20 sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids.

HCV amino acid sequences comprising epitopes can be identified in a number of ways. For example, the entire protein sequence corresponding to each of the NS5, envelope 1, and core regions can be screened by preparing a series of short peptides that together span

25

- 31 -

the entire protein sequence of such regions. By starting with, for example, peptides of approximately 100 amino acids, it would be routine to test each peptide for the presence of epitope(s) showing a desired reactivity, and then testing progressively smaller and overlapping fragments from an identified peptides of 100 amino acids to map the epitope of interest. Screening such peptides in an immunoassay is within the skill of the art. It is also known to carry out a computer analysis of a protein sequence to identify potential epitopes, and then prepare peptides comprising the identified regions for screening.

5

10

The immunogenicity of the epitopes of HCV may also be enhanced by preparing them in mammalian or yeast 15 systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B surface antigen. See, e.g., US 4,722,840. Constructs wherein the HCV epitope is linked directly to the particle-forming protein coding sequences 20 produce hybrids which are immunogenic with respect to the HCV epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles constructed from particle forming protein which include 25 HCV sequences are immunogenic with respect to HCV and HBV.

- 32 -

Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in S. cerevisiae (P. Valenzuela et al. (1982)), as well as in, for example, mammalian cells (P. Valenzuela et al. 1984)). The formation of such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath et al. (1984). Constructs of the pre-S-HBSAg particle expressible in yeast are disclosed in EPO 174,444, published March 19, 1986; hybrids including heterologous viral sequences for yeast expression are disclosed in EPO 175,261, published March 26, 1966. These constructs may also be expressed in mammalian cells such as Chinese hamster ovary (CHO) cells using an SV40-dihydrofolate reductase vector (Michelle et al. (1984)).

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an HCV epitope. In this replacement, regions which are not required to mediate the aggregation of the units to form immunogenic particles in yeast of mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the HCV epitope.

25

•

10

15

20

### Vaccines

Vaccines may be prepared from one or more

- 33 -

immunogenic peptides derived from HCV. The observed homology between HCV and Flaviviruses provides information concerning the peptides which are likely to be most effective as vaccines, as well as the regions of the genome in which they are encoded.

5

10

15

20

25

Multivalent vaccines against HCV may be comprised of one or more epitopes from one or more proteins derived from the NS5, envelope 1, and core regions. In particular, vaccines are contemplated comprising one or more HCV proteins or subunit antigens derived from the NS5, envelope 1, and core regions. The 5'UT region is not known to be translated.

The preparation of vaccines which contain an immunogenic peptide as an active ingredient, is known to one skilled in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or

- 34 -

emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-theronyl-D- isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl- D-isoglutamine (CGP 11637, referred to as nor-MDP), N- acetylmuramyl-Lalanyl-D-isoglutaminyl-L-alanine-2-(1- 2-dipalmitoyl -sn-glycero-3-hydroxyphosphoryloxy)- ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which 10 contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies 15 directed against an immunogenic peptide containing an HCV antigenic sequence resulting from administration of this peptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such

20

25

- 35 -

suppositories may be formed from mixtures containing the active ingredient in the range of 0/5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

The examples below are provided for illustrative purposes and are not intended to limit the scope of the present invention.

#### I. Detection of HCV RNA from Serum

10

RNA was extracted from serum using guanidinium salt, phenol and chloroform according to the

15 instructions of the kit manufacturer (RNAzol B kit, Cinna/Biotecx). Extracted RNA was precipitated with isopropanol and washed with ethanol. A total of 25 µl serum was processed for RNA isolation, and the purified RNA was resuspended in 5 µl diethyl

20 pyrocarbonate treated water for subsequent cDNA synthesis.

## II. <u>cDNA Synthesis and Polymerase Chain Reaction (PCR)</u> Amplification

25 Table 1 lists the sequence and position (with reference to HCV1) of all the PCR primers and probes used in these examples. Letter designations for

nucleotides are consistent with 37 C.F.R. \$\$1.821-1.825. Thus, the letters A, C, G, T, and U are used in the ordinary sense of adenine, cytosine, guanine, thymine, and uracil. The letter M means A or C; R 5 means A or G; W means A or T/U; S means C or G; Y means C or T/U; K means G or T/U; V means A or C or G, not T/U; H means A or C or T/U, not G; D means A or G or T/U, not C; B means C or G or T/U, not A; N means (A or C or G or T/U) or (unknown or other). Table 1 is set forth below:

Table 1

10

10

15

20

25

. ; ;

For cDNA synthesis and PCR amplification, a protocol developed by Perkin-Elmer/Cetus (GeneAmp® RNA PCR kit) was used. Both random hexamer and primers with specific complementary sequences to HCV were employed to prime the reverse transcription (RT) reaction. All processes, except for adding and mixing reaction components, were performed in a thermal cycler (MJ Research, Inc.). The first strand cDNA synthesis reaction was inactivated at 99°C for 5 min, and then cooled at 50°C for 5 min before adding reaction components for subsequent amplification. After an initial 5 cycles of 97°C for 1 min, 50°C for 2 min, and 72°C for 3 min, 30 cycles of 94°C for 1 min, 55°C for 2 min, and 72°C for 3 min followed, and then a final 7 min of elongation at 72°C.

For the genotyping analysis, sequences 67 and 68 were used as primers in the PCR reaction. These primers amplify a segment corresponding to the core and envelope regions. After amplification, the reaction products were separated on an agarose gel and then transferred to a nylon membrane. The immobilized reaction products were allowed to hybridize with a 32p-labelled nucleic acid corresponding to either Genotype I (core or envelope 1) or Genotype II (core or envelope 1). Nucleic acid corresponding to Genotype 1 comprised sequences numbered 69 (core), 71 (envelope), and 73 (envelope). Nucleic acid corresponding to

- 38 -

Genotype II comprised sequences numbered 70 (core), 72 (envelope), and 74 (envelope).

The Genotype I probes only hybridized to the product amplified from isolates which had Genotype I sequence. Similarly, Genotype II probes only hybridized to the product amplified from isolates which had Genotype II sequence.

In another experiment, PCR products were generated using sequences 79 and 80. The products were analyzed as described above except Sequence No. 73 was used to detect Genotype I, Sequence No. 74 was used to detect Genotype II, Sequence No. 77 (5'UT) was used to detect Genotype III, and Sequence No. 78 (5'UT) was used to detect Genotype IV. Each sequence hybridized in a genotype specific manner.

# III. Detection of HCV GI-GIV using a sandwich hybridization assay for HCV RNA

10

15

20

25

An amplified solution phase nucleic acid sandwich hybridization assay format is described in this example. The assay format employs several nucleic acid probes to effect capture and detection. A capture probe nucleic acid is capable of associating a complementary probe bound to a solid support and HCV nucleic acid to effect capture. A detection probe nucleic acid has a first segment (A) that binds to HCV nucleic acid and a second segment (B) that hybridizes to a second amplifier nucleic acid.

The amplifier nucleic acid has a first segment (B\*)
that hybridizes to segment (B) of the probe nucleic
acid and also comprises fifteen iterations of a segment
(C). Segment C of the amplifier nucleic acid is
capable of hybridizing to three labeled nucleic acids.

Nucleic acid sequences which correspond to nucleotide sequences of the envelope 1 gene of Group I HCV isolates are set forth in sequences numbered 81-99. Table 2 sets forth the area of the HCV genome to which the nucleic acid sequences correspond and a preferred use of the sequences.

10

Table 2

	Probe Type	Sequence No.	Complement of
15			Nucleotide Numbers
	<del></del>		
	Label	81	879-911
	Label	82	912-944
	Capture	83	945-977
20	Label	84	978-1010
	Label	85	1011-1043
	Label	86	1044-1076
	Label	87	1077-1109
	Capture	88	1110-1142
25	Label	89	1143-1175

- 40 -

Table 2 continued

	Probe Type	Sequence No.	Complement of Nucleotide Numbers
5	C223E3=2EE=2EE	:5595#25##P6#555###	1176-1208
	Label	90	<del>-</del> :
	Label	91	1209-1241
	Label	92	1242=1274
		93	1275-1307
10	Capture	94	1308-1340
	Label	-V	1341-1373
	Label	95	1374-1406
	. Label	· 96	1407-1439
	Label	97	
	Capture	98	1440-1472
15	Label	99	1473-1505

Nucleic acid sequences which correspond to nucleotide sequences of the envelope 1 gene of Group II HCV isolates are set forth in sequences 100-118. Table 3 sets forth the area of the HCV genome to which the nucleic acid corresponds and the preferred use of the sequences.

20

- 41 -

Table 3

	Probe Type	Sequence No.	Complement of Nucleotide Numbers
5	Label	100	879-911
	Label	101	912-944
	Capture	102	945-977
	Label	103	978-1010
10	Label	104	1011-1043
	Label	105	1044-1076
	Label	106	1077-1109
	Capture	107	1110-1142
	Label	108	1143-1175
15	Label	109	1176-1208
	Label	110	1209-1241
	Label	111	1242=1274
	Capture	112	1275-1307
	Label	113	1308-1340
20	Label	114	1341-1373
	Label	115	1374-1406
	Label	116	1407-1439
	Capture	117	1440-1472
	Label	118	1473-1505
25			

Nucleic acid sequences which correspond to nucleotide sequences in the C gene and the 5'UT region

- 42 -

are set forth in sequences 119-145. Table 4 identifies the sequence with a preferred use.

Table 4

5		
3	Probe Type	Sequence No.
	Capture	119
	Label	120
10	Label	121
,	Label	122
	Capture	123
	Label	124
	Label	125
15	Label	126
	Capture	127
	Label	128
	Label	129
	Label	130
20	Capture	131
	Label	132
	Label	133
	Label	134
	Label	135
25	Capture	136
	Label	137
	Label	138

- 43 -

Table 4 continued

	Probe Type	Sequence No.
5	Label	139
	Capture	140
	Label	141
	Label	142
	Label	143
10	Capture	144
	Label	145

The detection and capture probe HCV-specific segments, and their respective names as used in this assay were as follows.

Capture sequences are sequences numbered 119-122 and 141-144.

Detection sequences are sequences numbered 119-140.

the sequences substantially complementary to the HCV sequences, a 5' extension (B) which extension (B) is complementary to a segment of the second amplifier nucleic acid. The extension (B) sequence is identified in the Sequence Listing as Sequence No. 146, and is reproduced below.

**AGGCATAGGACCCGTGTCTT** 

- 44 -

the sequences substantially complementary to HCV sequences, a sequence complementary to DNA bound to a solid phase. The sequence complementary to DNA bound to a solid support was carried downstream from the capture sequence. The sequence complementary to the DNA bound to the support is set forth as Sequence No. 147 and is reproduced below.

#### CTTCTTTGGAGAAAGTGGTG

25

Microtiter plates were prepared as follows. White Microlite 1 Removawell strips (polystyrene microtiter plates, 96 wells/plate) were purchased from Dynatech Inc.

Each well was filled with 200 μl 1 N HCl and incubated at room temperature for 15-20 min. The plates were then washed 4 times with 1X PBS and the wells aspirated to remove liquid. The wells were then filled with 200 μl 1 N NaOH and incubated at room temperature for 15-20 min. The plates were again washed 4 times with 1X PBS and the wells aspirated to remove liquid.

Poly(phe-lys) was purchased from Sigma Chemicals, Inc. This polypeptide has a 1:1 molar ratio of phe:lys and an average m.w. of 47,900 gm/mole. It has an average length of 309 amino acids and contains 155 amines/mole. A 1 mg/ml solution of the polypeptide was mixed with 2M NaCl/lX PBS to a final concentration of

0.1 mg/ml (pH 6.0). A volume of 200  $\mu$ l of this solution was added to each well. The plate was wrapped in plastic to prevent drying and incubated at 30°C overnight. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid.

5

10

15

The following procedure was used to couple the nucleic acid, a complementary sequence to Sequence No. 147, to the plates, hereinafter referred to as immobilized nucleic acid. Synthesis of immobilized nucleic acid having a sequence complementary to Sequence No. 133 was described in EPA 883096976. A quantity of 20 mg disuccinimidyl suberate was dissolved in 300  $\mu$ l dimethyl formamide (DMF). A quantity of 26 OD<sub>260</sub> units of immobilized nucleic acid was added to 100  $\mu$ l coupling buffer (50 mM sodium phosphate, pH 7.8). The coupling mixture was then added to the DSS-DMF solution and stirred with a magnetic stirrer for 30 min. An NAP-25 column was equilibrated with 10 mM sodium phosphate, pH 6.5. The coupling mixture DSS-DMF solution was added to 2 ml 10 mM sodium 20 phosphate, pH 6.5, at 4°C. The mixture was vortexed to mix and loaded onto the equilibrated NAP-25 column. DSS-activated immobilized nucleic acid DNA was eluted from the column with 3.5 ml 10 mM sodium phosphate, pH 6.5. A quantity of 5.6 OD 260 units of eluted 25 DSS-activated immobilized nucleic acid DNA was added to 1500 ml 50 mM sodium phosphate, pH 7.8. A volume of 50

- 46 -

µl of this solution was added to each well and the plates were incubated overnight. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid.

5

10

Final stripping of plates was accomplished as follows. A volume of 200  $\mu$ l of 0.2N NaOH containing 0.5% (w/v) SDS was added to each well. The plate was wrapped in plastic and incubated at 65°C for 60 min. The plate was then washed 4 times with 1X PBS and the wells aspirated to remove liquid. The stripped plate was stored with desiccant beads at 2-8°C.

Serum samples to be assayed were analyzed using PCR followed by sequence analysis to determine the genotype.

Sample preparation consisted of delivering 50 μl of the serum sample and 150 μl P-K Buffer (2 mg/ml proteinase K in 53 mM Tris-HCl, pH 8.0/0.6 M NaCl/0.06 M sodium citrate/8 mM EDTA, pH 8.0/1.3%SDS/16μg/ml sonicated salmon sperm DNA/7% formamide/50 fmoles capture probes/160 fmoles detection probes) to each well. Plates were agitated to mix the contents in the well, covered and incubated for 16 hr at 62°C.

After a further 10 minute period at room temperature, the contents of each well were aspirated to remove all fluid, and the wells washed 2X with washing buffer (0.1% SDS/0.015 M NaCl/ 0.0015 M sodium citrate). The amplifier nucleic acid was then added to

each well (50 µl of 0.7 fmole/µl solution in 0..48 M NaCl/0.048 M sodium citrate/0.1% SDS/0.5% "blocking reagent" (Boehringer Mannheim, catalog No. 1096 176)). After covering the plates and agitating to mix the contents in the wells, the plates were incubated for 30 min. at 52°C.

After a further 10 min period at room temperature, the wells were washed as described above.

Alkaline phosphatase label nucleic acid, disclosed in EP 883096976, was then added to each well (50 µl/well of 2.66 fmoles/µl). After incubation at 52°C for 15 min., and 10 min. at room temperature, the wells were washed twice as above and then 3X with 0.015 M NaCl/0.0015 M sodium citrate.

10

25

An enzyme-triggered dioxetane (Schaap et al., Tet. Lett. (1987) 28:1159-1162 and EPA Pub. No. 0254051), obtained from Lumigen, Inc., was employed. A quantity of 50 μl Lumiphos 530 (Lumigen) was added to each well. The wells were tapped lightly so that the reagent would fall to the bottom and gently swirled to distribute the reagent evenly over the bottom. The wells were covered and incubated at 37°C for 20-40 min.

Plates were then read on a Dynatech ML 1000 luminometer. Output was given as the full integral of the light produced during the reaction.

The assay positively detected each of the serum samples, regardless of genotype.

PCT/US92/04036 WO 92/19743

- 48 -

### IV. Expression of the Polypeptide Encoded in Sequences Defined by Differing Genotypes

HCV polypeptides encoded by a sequence within sequences 1-66 are expressed as a fusion polypeptide with superoxide dismutase (SOD). A cDNA carrying such sequences is subcloned into the expression vector psoDcfl (Steimer et al. 1986)).

First, DNA isolated from pSODcfl is treated with BamHI and EcoRI, and the following linker was ligated into the linear DNA created by the restriction enzymes:

GAT CCT GGA ATT CTG ATA AGA

CCT TAA GAC TAT TTT AA

After cloning, the plasmid containing the insert is

Plasmid containing the insert is restricted with isolated. EcoRI. The HCV cDNA is ligated into this EcoRI 15 linearized plasmid DNA. The DNA mixture is used to transform E. coli strain D1210 (Sadler et al. (1980)). Polypeptides are isolated on gels.

20

25

10

### Antigenicity of Polypeptides

The antigenicity of polypeptides formed in Section ٧. IV is evaluated in the following manner. Polyethylene pins arranged on a block in an 8 12 array (Coselco Mimetopes, Victoria, Australia) are prepared by placing the pins in a bath (20% v/v piperidine in dimethylformamide (DMF)) for 30 minutes at room

10

15

temperature. The pins are removed, washed in DMF for 5 minutes, then washed in methanol four times (2 min/wash). The pins are allowed to air dry for at least 10 minutes, then washed a final time in DMF (5Min). 1-Hydroxybenzotriazole (HOBt, 367 mg) is dissolved in DMF (80  $\mu$ L) for use in coupling Fmoc-protected polypeptides prepared in Section IV.

The protected amino acids are placed in micro-titer plate wells with HOBt, and the pin block placed over the plate, immersing the pins in the wells. The assembly is then sealed in a plastic bag and allowed to react at 25°C for 18 hours to couple the first amino acids to the pins. The block is then removed, and the pins washed with DMF (2 min.), MeOH (4 x, 2 min.), and again with DMF (2 min.) to clean and deprotect the bound amino acids. The procedure is repeated for each additional amino acid coupled, until all octamers are prepared.

The free N-termini are then acetylated to

compensate for the free amide, as most of the epitopes
are not found at the N-terminus and thus would not have
the associated positive charge. Acetylation is
accomplished by filling the wells of a microtiter plate
with DMF/acetic anhydride/triethylamine (5:2:1 v/v/v)

and allowing the pins to react in the wells for 90
minutes at 20°C. The pins are then washed with DMF (2)

10

15

20

25

min.) and MeOH (4 x, 2 min.), and air dried for at least 10 minutes.

The side chain protecting groups are removed by treating the pins with trifluoroacetic acid/phenol/dithioethane (95:2.5:1.5, v/v/v) in polypropylene bags for 4 hours at room temperature. The pins are then washed in dichloromethane (2 x, 2 min.), 5% di-isopropylethylamine/dichloromethane (2 x, 5 min.), dichloromethane (5 min.), and air-dried for at least 10 minutes. The pins are then washed in water (2 min.), MeOH (18 hours), dried in vacuo, and stored in sealed plastic bags over silica gel. IV.B.15.b Assay of Peptides.

Octamer-bearing pins are treated by sonicating for 30 minutes in a disruption buffer (1% sodium dodecylsulfate, 0.1% 2-mercaptoethanol, 0.1 M NaH2PO4) at 60°C. The pins are then immersed several times in water (60°C), followed by boiling MeOH (2 min.), and allowed to air dry.

The pins are then precoated for 1 hour at 25°C in microtiter wells containing 200 µL blocking buffer (1% ovalbumin, 1% BSA, 0.1% Tween, and 0.05% NaN3 in PBS), with agitation. The pins are then immersed in microtiter wells containing 175 µL antisera obtained from human patients diagnosed as having HCV and allowed to incubate at 4°C overnight. The formation of a complex between polyclonal antibodies of the serum and

- 51 -

the polypeptide initiates that the peptides give rise to an immune response in vivo. Such peptides are candidates for the development of vaccines.

Thus, this invention has been described and illustrated. It will be apparent to those skilled in the art that many variations and modifications can be made without departing from the purview of the appended claims and without departing from the teaching and scope of the present invention.

- 52 -

#### SEQUENCE LISTING

### (1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Tai-An Cha
  - (ii) TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS
- 10 (iii) NUMBER OF SEQUENCES: 147
  - (iv) CORRESPONDENCE ADDRESS:
    - (A) ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
    - (B) STREET: 600 Atlantic Avenue
- 15 (C) CITY: Boston
  - (D) STATE: Massachusetts
  - (E) COUNTRY: USA
  - (F) ZIP: 02210
- 20 (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Diskette, 5.25 inch
  - (B) COMPUTER: IEM compatible
  - (C) OPERATING SYSTEM: MS-DOS Version 3.3
  - (D) SOFTWARE: WordPerfect 5.1

- 53 -

(vi) CURRENT APPLICATION DATA: (A) APPLICATION NUMBER: Not Available (B) FILING DATE: Not Available (C) CLASSIFICATION: Not Available 5 (vii) PRIOR APPLICATION DATA: (A) APPLICATION NUMBER: 07/697,326 (B) FILING DATE: 8 May 1991 10 (viii) ATTORNEY/AGENT INFORMATION: (A) NAME: Janiuk, Anthony J. (B) REGISTRATION NUMBER: 29,809 (C) REFERENCE/DOCKET NUMBER: C0772/7000 15 (ix) TELECOMMUNICATION INFORMATION: (A) TELEPHONE: (617) 720-3500 (B) TELEFAX: (617) 720-2441 (C) TELEX: EZEKIEL 20 (2) INFORMATION FOR SEQ ID NO: 1: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 340 nucleotides (B) TYPE: nucleic acid (C) STRANDEDNESS: single 25 (D) TOPOLOGY: linear

- 54 -

(ii) MOLECULE TYPE: DNA

- 55 -

	•	(vi) ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: ns5i21	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2	
5		CTCCACAGTC ACTGAGAGCG ACATCCGTAC GGAGGAGGCA	40
5		ATTTACCAAT GTTGTGACCT GGACCCCCAA GCCCGCATGG	80
		CCATCAAGTC CCTCACTGAG AGGCTTTATG TCGGGGGCCC	120
			160
			200
LO			240
LU			280
			320
			340
15	(2)	INFORMATION FOR SEQ ID NO: 3:	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
20		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
25		(vi) ORIGINAL SOURCE:	
		(C) individual isolate: ns5ptl	

		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3	
		CTCCACAGTC ACTGAGAGCG ACATCCGTAC GGAGGAGGCA	40
		ATCTACCAAT GTTGTGATCT GGACCCCCAA GCCCGCGTGG	80
		ATCTACCAAT GTTGTGATCT GGACCCCCCCCCCCCCCC	120
		CCATCAAGTC CCTCACTGAG AGGCTTACG CTACCGCAGG	160
5		TCTTACCAAT TCAAGGGGGG AGAACTGCGG CTACCGCAGG	200
		TGCCGGGCGA GCGGCGTACT GACAACTAGC TGTGGTAATA	240
		CCCTCACTTG CTACATCAAG GCCCGGGCAG CCTGTCGAGC	280
		CGCAGGGCTC CGGGACTGCA CCATGCTCGT GTGTGGTGAC	320
		GACTTGGTCG TTATCTGTGA GAGTGCGGGG GTCCAGGAGG	
10		ACGCGGCGAG CCTGAGAGCC	340
	(2)	INFORMATION FOR SEQ ID NO: 4	
		(i) SEQUENCE CHARACTERISTICS:	
15		(A) LENGTH: 340 nucleotides	
15		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
20		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: IISOGAZ	
05		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4	
25		CTCTACAGTC ACTGAGAACG ACATCCGTAC GGAGGAGGCA	40
		AMERICANT GTTGTGACCT GGACCCCCAA GCCCGCGTGG	80

- 57 -

5	CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC 120  CCTTACCAAT TCAAGGGGGG AAAACTGCGG CTATCGCAGG 160  TGCCGCGCGA GCGGCGTACT GACAACTAGC TGTGGTAACA 200  CCCTCACTTG CTACATTAAG GCCCGGGCAG CCTGTCGAGC 240  CGCAGGGCTC CAGGACTGCA CCATGCTCGT GTGTGGCGAC 280  GACTTAGTCG TTATCTGTGA GAGTGCGGGA GTCCAGGAGG 320  ACGCGGCGAA CTTGAGAGCC 340
10	INFORMATION FOR SEQ ID NO: 5  (i) SEQUENCE CHARACTERISTICS:
15	(A) LENGTH: 340 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear  (ii) MOLECULE TYPE: DNA
20	(vi) ORIGINAL SOURCE:  (C) INDIVIDUAL ISOLATE: ns5us17
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5  CTCCACAGTC ACTGAGAGCG ATATCCGTAC GGAGGAGGCA 40  ATCTACCAGT GTTGTGACCT GGACCCCCAA GCCCGCGTGG 80  CCATCAAGTC CCTCACCGAG AGGCTTTATG TCGGGGGCCC 120  TCTTACCAAT TCAAGGGGGG AAAACTGCGG CTATCGCAGG 160  TGCCGCGCAA GCGGCGTACT GACAACTAGC TGTGGTAACA 200

- 58 -

		CCCTCACTTG TTACATCAAG GCCCAAGCAG CCTGTCGAGC	240 280
		CGCAGGGCTC CGGGACTGCA CCATGCTCGT GTGTGGCGAC	
		GACTTAGTCG TTATCTGTGA AAGTCAGGGA GTCCAGGAGG	320
		ATGCAGCGAA CCTGAGAGCC	340
		ATGCAGCGAA CCIGAGGGG	
5	(2)	INFORMATION FOR SEQ ID NO: 6	
	•	(i) SEQUENCE CHARACTERISTICS:	
·· ·		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(8)	
		(ii) MOLECULE TYPE: DNA	
15		<pre>(vi) ORIGINAL SOURCE:    (C) INDIVIDUAL ISOLATE: ns5sp2</pre>	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6	
		CTCTACAGTC ACTGAGAGCG ATATCCGTAC GGAGGAGGCA	40
20		ATCTACCAAT GTTGTGACCT GGACCCCGAA GCCCGTGTGG	80
		CCATCAAGTC CCTCACTGAG AGGCTTTATG TTGGGGGCCC	120
		TCTTACCAAT TCAAGGGGGG AGAACTGCGG CTACCGCAGG	160
		TGCCGCGCAA GCGGCGTACT GACGACTAGC TGTGGTAATA	200
		TGCCGCGCAA GCGGCGTACT ONOCCOCAGG CCTGTCGAGC CCCTCACTTG TTACATCAAG GCCCGGGCAG CCTGTCGAGC	240
25		CCCTCACTTG TTACATCARG GCCCCCCTCGT GTGTGGCGAC	280
		CELVERIAL CARREST CONTRACTOR	

- 59 -

		GACCTAGTCG TTATCTGCGA AAGTGCGGGG GTCCAGGAGG	320
		ACGCGGCGAG CCTGAGAGCC	340
		ACGCGGCGAG CCTGAGAGGG	
	(2)	INFORMATION FOR SEQ ID NO: 7	
5		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
10		(D) TOPOLOGY: linear	
10			
		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	
15		(C) INDIVIDUAL ISOLATE: ns5j1	٠
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7	
		CTCCACAGTC ACTGAGAATG ACACCCGTGT TGAGGAGTCA	40
		ATTTACCAAT GTTGTGACTT GGCCCCCGAA GCCAGACAGG	80
		CCATAGGTC GCTCACAGAG CGGCTCTATG TCGGGGGTCC	120
20		TATGACTAAC TCCAAAGGGC AGAACTGCGG CTATCGCCGG	160
		TATGACTAAC TCCAAAGGGC AGAACIGCGG CIAICGGTAATA	20
		TGCCGCGCGA GCGGCGTGCT GACGACTAGC TGCGGTAATA	24
		CCCTCACATG CTACCTGAAG GCCACAGCGG CCTGTCGAGC	28
		TGCCAAGCTC CAGGACTGCA CGATGCTCGT GAACGGAGAC	32
25		GACCTTGTCG TTATCTGTGA AAGCGCGGGG AACCAAGAGG	34
		> CCCCCC > CCCT CCACCC	J#

- 60 -

	(2)	INFORMATION FOR SEQ ID NO: 8	
5		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 340 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
10		<pre>(vi) ORIGINAL SOURCE:    (C) INDIVIDUAL ISOLATE: ns5kl</pre>	
15		CTCAACGGTC ACTGAGAATG ACATCCGTGT TGAGGAGICA CTCAACGGTC ACTGAGAATG ACATCCGTGT TGAGGAGICA ATTTACCAAA GTTGTGACTT GGCCCCCGAG GCCAGACAAG CCATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGCCC 1 CCATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGCCC 1	40 80 .20
20	)	CCTGACTAAT TCAAAAGGGC AGAACTOOO  TGCCGCGCCA GCGGTGTGCT GACGACTAGC TGCGGTAATA  TGCCGCGCCA GCGGTGTGCT GACGACTGCGG CCTGTAGAGC  CCCTCACATG TTACTTGAAG GCCACTGCGG CCTGTAGAGC	200 240 280 320 340
2	5 (	(2) INFORMATION FOR SEQ ID NO: 9	

- 61 -

	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 340 nucleotides
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: single
5	(D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA
••	(vi) ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: DS5kl.1
10	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9
	COCARCECTO ACCEAGAATE ACATCCETET TEAGGACTO.
	ATTUATCAAT GTTGTGCCTT GGCCCCCGAG GCIAGAGIO
15	COMPAGGTC GCTCACAGAG CGGCTTTATA ICGGGGGGGGG
	COTCACCAAT TCAAAGGGGC AGAACTGCGG TIATCGGGG
	TECCECCIA GEGGEGIACI GAEGACEAGE IGEGGIALITA
	COUTTACATG TTACTTGAAG GCCTCTGCAG CCIGICGAG
	COCCALCACT CAGGACTGCA CGATGCTCGT GIGIGOCOLO
20	GACCTTGTCG TTATCTGTGA AAGCGCGGGA ACCCAGGAGA
	ACGCGGCGAA CCTACGAGTC
	(2) INFORMATION FOR SEQ ID NO: 10
25	<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 340 nucleotides</li><li>(B) TYPE: nucleic acid</li></ul>

*:*;;

- 62 -

		(C)	STRANDEDNESS: single	
		(D)	TOPOLOGY: linear	
5		(ii) MOI	ECULE TYPE: DNA	
3			GINAL SOURCE: INDIVIDUAL ISOLATE: ns5gh6	
		(xi) SEC	QUENCE DESCRIPTION: SEQ ID NO: 10	
10		CTCAACGGTC	ACTGAGAGTG ACATCCGTGT CGAGGAGTCG	40
		ATTTACCAAT	GTTGTGACTT GGCCCCCGAA GCCAGGCAGG	80
		CCATAAGGT	GCTCACCGAG CGACTTTATA TCGGGGGCCC	120
			TCAAAAGGGC AGAACTGCGG TTATCGCCGG	160
			A GCGGCGTGCT GACGACTAGC TGCGGTAATA	200
15		CCCTCACATO	TTACTTGAAG GCCTCTGCAG CCTGTCGAGC	240
			CAGGACTGCA CGATGCTCGT GAACGGGGAC	
		GACCTTGTC	TTATCTGCGA GAGCGCGGGA ACCCAAGAGG	320
			CCTACGAGTC	340
20	(2)	INFORMATIO	ON FOR SEQ ID NO: 11	
			QUENCE CHARACTERISTICS:	
			LENGTH: 340 nucleotides	
			TYPE: nucleic acid	
25		(C)	STRANDEDNESS: single	
		(D)	TOPOLOGY: linear	
			•	

- 63 -

		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: ns5spl	
5			SEQUENCE DESCRIPTION: SEQ ID NO: 11	
		(xi)	AGTC ACTGAGAGTG ACATCCGTGT TGAGGAGTCA	40
		CTCCAC	CAAT GTTGTGACTT GGCCCCCGAA GCCAGACAGG	80
		ATTTAC	CAAT GTTGTGACTT GGCCCCCGAA GCCAGACACC	120
	:	CTATAA	GGTC GCTCACAGAG CGGCTGTACA TCGGGGGTCC	160
10		CCTGAC'	TAAT TCAAAAGGGC AGAACTGCGG CTATCGCCGG	200
		TGCCGC	GCAA GCGGCGTGCT GACGACTAGC TGCGGTAACA	
			ACATG TTACTTGAAG GCCTCTGCGG CCTGTCGAGC	-
		TGCGAA	AGCTC CAGGACTGCA CGATGCTCGT GTGCGGTGAC	
			GTCG TTATCTGTGA GAGCGCGGGA ACCCAAGAGG	340
15		ACGCGG	GCGAG CCTACGAGTC	340
	(2)	INFORM	MATION FOR SEQ ID NO: 12	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 340 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	

- 64 -

	(C) individual isolate: ns5sp3	
5	(XI) SEQUENCE DESCRIPTION: SEQ ID NO: 12  CTCAACAGTC ACTGAGAGTG ACATCCGTGT TGAGGAGTCA  ATCTACCAAT GTTGTGACTT GGCCCCCGAA GCCAGACAGG  CTATAAGGTC GCTCACAGAG CGGCTTTACA TCGGGGGTCC  CCTGACTAAT TCAAAAGGGC AGAACTGCGG CTATCGCCGG  TGCCGCGCAA GCGGCGTGCT GACGACTAGC TGCGGTAATA	40 80 120 160 200
10	CCCTCACATG TTACCTGAAG GCCAGTGCGG CCTGTCGAGC TGCGAAGCTC CAGGACTGCA CAATGCTCGT GTGCGGTGAC GACCTTGTCG TTATCTGTGA GAGCGCGGGG ACCCAAGAGG ACGCGGCGAG CCTACGAGTC	28 32 34
(2) 15	INFORMATION FOR SEQ ID NO: 13  (i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 340 nucleotides  (B) TYPE: nucleic acid	
20	(C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA	
25	<pre>(vi) ORIGINAL SOURCE:</pre>	



**- 65 -**

		CTCAACCGTC ACTGAGAGAG ACATCAGAAC TGAGGAGTCC	4.
		ATATACCGAG CCTGCTCCCT GCCTGAGGAG GCTCACATTG	80
		CCATACACTC GCTGACTGAG AGGCTCTACG TGGGAGGGCC	120
		CATGTTCAAC AGCAAGGGCC AGACCTGCGG GTACAGGCGT	160
5		TGCCGCGCCA GCGGGGTGCT CACCACTAGC ATGGGGAACA	200
		CCATCACATG CTATGTAAAA GCCCTAGCGG CTTGCAAGGC	240
		TGCAGGGATA GTTGCACCCT CAATGCTGGT ATGCGGCGAC	280
		GACTTAGTTG TCATCTCAGA AAGCCAGGGG ACTGAGGAGG	320
		ACGAGCGGAA CCTGAGAGCT	340
10			
	(2)	INFORMATION FOR SEQ ID NO: 14	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
15		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20		(II) MODECODE IIFE. DNA	
20		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5arg8	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14	
25		CTCTACAGTC ACGTAAAAGG ACATCACATC CTAGGAGTCC	40
43		ATCTACCAGT CCTGTTCACT GCCGAGGAG GCTCGAACTG	80
		CTATACACTC ACTGACTGAG AGACTATACG TAGGGGGGCC	120
		CIVINCUCIC VCIQUCIQUO VQVCIVIVCQ IVQQQQQCC	

- 66 -

		CATGACAAAC AGCAAGGGCC AATCCTGCGG GTACAGGCGT	160
		CCCACTAGE CACCACCAGE RICCOMME	200
		TGCCGCGCGA GCGCAGTOTO GCCAGGGCGG CGTGTAACGC CACTCACGTG CTACGTAAAA GCCAGGGCGG CGTGTAACGC	240
		CACTCACGTG CTACGTAGGT GTGCGGTGAC CGCGGGGATT GTTGCTCCCA CCATGCTGGT GTGCGGTGAC	280
		CGCGGGGATT GTTGCTCCCA COLIFORNIA GAGTCAAGGG GCTGAGGAGG GACCTGGTCG TCATCTCAGA GAGTCAAGGG GCTGAGGAGG	320
5		GACCTGGTCG TCATCTCAGA GACTC	340
		ACGAGCAGAA CCTGAGAGTC	
	(2)	INFORMATION FOR SEQ ID NO: 15	
	•	(i) SEQUENCE CHARACTERISTICS:	
10		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
15		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5i10	•
20		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15	
		(xi) SEQUENCE DESCRIPTION. DAG COAGGAGTCC CTCTACAGTC ACAGAGAGGG ACATCAGAAC CGAGGAGTCC	40
		ATCTATCTGT CCTGCTCACT GCCTGAGGAG GCCCGAACTG	80
		ATCTATCTGT CCTGCTCACT COCCUTATACG TAGGGGGGCC CTATACACTC ACTGACTGAG AGACTGTACG TAGGGGGGCC	120
		CTATACACTC ACTGACTORS THEOGRAPHIC CATGACAAAC AGCAAGGGGC AATCCTGCGG GTACAGGCGT CATGACAAAC AGCAAGGGGC AATCCTGCGCAACA	160
25		TGCCGCGCGA GCGGAGTGCT CACCACCAGC ATGGGCAACA	200
		TGCCGCGCGA GCGGAGTGCT CHOCKER CGCTCACGTG CTACGTGAAA GCCAGAGCGG CGTGTAACGC	240
		CGCTCACGTG CTACGTGATA COCTO	

- 67 -

		CGCGGGCATT GTTGCTCCCA CCATGTTGGT GTGCGGCGAC	280
		GACCTGGTTG TCATCTCAGA GAGTCAGGGG GTCGAGGAAG	320
		ATGAGCGGAA CCTGAGAGTC	340
5	(2)	INFORMATION FOR SEQ ID NO: 16	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
15		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5arg6	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16	
		CTCTACAGTC ACGGAGAGGG ACATCAGAAC CGAGGAGTCC	40
20		ATCTATCTGT CCTGTTCACT GCCTGAGGAG GCTCGAACTG	80
		CCATACACTC ACTGACTGAG AGGCTGTACG TAGGGGGGCC	120
		CATGACAAAC AGCAAAGGGC AATCCTGCGG GTACAGGCGT	160
		TGCCGCGCGA GCGGAGTGCT CACCACCAGC ATGGGTAACA	200
		CACTCACGTG CTACGTGAAA GCTAAAGCGG CATGTAACGC	240
25		CGCGGGCATT GTTGCCCCCA CCATGTTGGT GTGCGGCGAC	280
		GACCTAGTCG TCATCTCAGA GAGTCAAGGG GTCGAGGAGG	320
		ATGAGCGAAA CCTGAGAGCT	340

- 68 -

	(2)	INFORMATION FOR SEQ ID NO: 17	
5		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 340 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
10		(ii) MOLECULE TYPE: DNA	
•		<pre>(vi) ORIGINAL SOURCE:    (C) INDIVIDUAL ISOLATE: ns5k2b</pre>	
15		(Xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17 CTCAACCGTC ACGGAGAGGG ACATAAGAAC AGAAGAATCC ATATATCAGG GTTGTTCCCT GCCTCAGGAG GCTAGAACTG CTATCCACTC GCTCACTGAG AGACTCTACG TAGGAGGGCC CATGACAAAC AGCAAGGGAC AATCCTGCGG TTACAGGCGT	40 80 120
20		TGCCGCGCCA GCGGGGTCTT CACCACCAGC ATGGGGAATA CCATGACATG CTACATCAAA GCCCTTGCAG CGTGCAAAGC TGCAGGGATC GTGGACCCTA TCATGCTGGT GTGTGGAGAC GACCTGGTCG TCATCTCGGA GAGCGAAGGT AACGAGGAGG ACGAGCGAAA CCTGAGAGCT	200 240 280 320 340
25	(2)	INFORMATION FOR SEQ ID NO: 18	
		(i) SEQUENCE CHARACTERISTICS:	

- 69 -

		(A) LENGTH: 340 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
5			
		(ii) MOLECULE TYPE: DNA	
		•	
		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5sa283	
10			
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18	
		CTCGACCGTT ACCGAACATG ACATAATGAC TGAAGAGTCT	40
		ATTTACCAAT CATTGTACTT GCAGCCTGAG GCGCGTGTGG	80
			120
15			160
		TGCCGCGCCA GCGGCGTCTT CACCACTAGT ATGGGCAACA	200
		CCATGACGTG CTACATTAAG GCTTTAGCCT CCTGTAGAGC	240
		CGCAAAGCTC CAGGACTGCA CGCTCCTGGT GTGTGGTGAT	320
		GATAAAGCGA CCTGAGAGCC	340
20			
	(2)	INFORMATION FOR SEQ ID NO: 19	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
25		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	

		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: ns5sa156	
5				
			SEQUENCE DESCRIPTION: SEQ ID NO: 19	
			CGTT ACCGAACATG ACATAATGAC TGAAGAGTCC	40
			CAAT CATTGTACTT GCAGCCTGAG GCACGCGCGG	
		CAATAC	GGTC ACTCACCCAA CGCCTGTACT GTGGAGGCCC	
10		CATGTA	TAAC AGCAAGGGGC AACAATGTGG TTACCGTAGA	
		TGCCGC	GCCA GCGGCGTCTT CACCACCAGT ATGGGCAACA	200
		CCATGA	CGTG CTACATCAAG GCTTCAGCCG CCTGTAGAGC	240
		TGCAAA	GCTC CAGGACTGCA CGCTCCTGGT GTGTGGTGTG	280
		ACCTTG	GTGG CCATTTGCGA GAGCCAAGGG ACGCACGAGG	320
15			CCTG CCTGAGAGTC	340
	(2)	INFORM	ATION FOR SEQ ID NO: 20	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 340 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	

- 71 -

		(C) INDIVIDUAL ISOLATE: ns5il1	
5		(XI) SEQUENCE DESCRIPTION: SEQ ID NO: 20 CTCTACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG ATATACCAGT GCTGTAACCT TGAACCGGAG GCCAGGAAAG TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC TATGTTCAAC AGCAAGGGGG CCCAGTGTGG TTATCGCCGT TGCCGTGCTA GTGGAGTCCT GCCTACCAGC TTCGGCAACA CAATCACTTG TTACATCAAG GCTAGAGCGG CTTCGAAGGC	160 200 240
10		CGCAGGCCTC CGGAACCCGG ACTTTCTTGT CTGCGGAGAT GATCTGGTCG TGGTGGCTGA GAGTGATGGC GTCGACGAGG ATAGAGCAGC CCTGAGAGCC	280 320 340
15	(2)	<pre>INFORMATION FOR SEQ ID NO: 21  (i) SEQUENCE CHARACTERISTICS:     (A) LENGTH: 340 nucleotides     (B) TYPE: nucleic acid     (C) STRANDEDNESS: single</pre>	
20		(D) TOPOLOGY: linear  (ii) MOLECULE TYPE: DNA	
25		<pre>(vi) ORIGINAL SOURCE:      (C) INDIVIDUAL ISOLATE: ns5i4  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21</pre>	

		CTCGACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG	40
		ATATACCAAT GCTGTAACCT TGAACCGGAG GCCAGGAAAG	80
		TGATCTCCTC CCTCACGGAG CGGCTTTACT GCGGGGGCCC	120
		TATGTTCAAT AGCAAGGGGG CCCAGTGTGG TTATCGCCGT	160
_		TGCCGTGCTA GTGGAGTTCT GCCTACCAGC TTCGGCAACA	200
5		CAATCACTTG TTACATCAAG GCTAGAGCGG CTGCGAAGGC	240
		CGCAGGGCTC CGGACCCCGG ACTTTCTCGT CTGCGGAGAT	280
	•	GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCGACGAGG	320
			340
		ATAGAACAGC CCTGCGAGCC	
10			
	(2)	INFORMATION FOR SEQ ID NO: 22	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 340 nucleotides	
15		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20		120	
20		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: ns5gh8	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22	
		CTCAACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG	40
25		ATATACCAAT GCTGTAACCT TGAACCGGAG GCCAGGAAAG	80
		TGATCTCCTC CCTCACGGAA CGGCTTTACT GCGGGGGCCC	120
		TOTAL CONTRACT OF THE PROPERTY	

	TATGTTCAAC AGCAAGGGGG CCCAGTGTGG TTATCGCCGT 160 TGCCGTGCCA GTGGAGTTCT GCCTACCAGC TTCGGCAACA 200 CAATCACTTG TTACATCAAA GCTAGAGCGG CTGCCGAAGC 240 CGCAGGCCTC CGGAACCCGG ACTTTCTTGT CTGCGGAGAT 280	
5	GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCATTOTTO 340 ATAGAGCAGC CCTGGGAGCC	
(2)	INFORMATION FOR SEQ ID NO: 23	
10	<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 100 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
15	(ii) MOLECULE TYPE: DNA	
	<pre>(vi) ORIGINAL SOURCE: (ATCC # 40394) (C) INDIVIDUAL ISOLATE: hcvl</pre>	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23  GACGGCGTTG GTAATGGCTC AGCTGCTCCG GATCCCACAA 40  GCCATCTTGG ACATGATCGC TGGTGCTCAC TGGGGAGTCC 80  TGGCGGGCAT AGCGTATTTC	)
25 (	2) INFORMATION FOR SEQ ID NO: 24	

		(i)	SEQUENCE CHARACTERISTICS:
			(A) LENGTH: 100 nucleotides
			(B) TYPE: nucleic acid
			(C) STRANDEDNESS: single
5			(D) TOPOLOGY: linear
		(ii)	MOLECULE TYPE: DNA
		(vi)	ORIGINAL SOURCE:
10			(C) INDIVIDUAL ISOLATE: US5
		(i \	SEQUENCE DESCRIPTION: SEQ ID NO: 24
			GTTG GTGGTAGCTC AGGTACTCCG GATCCCACAA 40
			ATGG ACATGATCGC TGGAGCCCAC TGGGGAGTCC 80
15		TGGCGG	GCAT AGCGTATTTC 100
	(0)	TYPODM	ATION FOR SEQ ID NO: 25
	(2)	INFORM	ATION FOR BEG 15 No. 25
		(i)	SEQUENCE CHARACTERISTICS:
20			(A) LENGTH: 100 nucleotides
			(B) TYPE: nucleic acid
			(C) STRANDEDNESS: single
			(D) TOPOLOGY: linear
25		(ii)	MOLECULE TYPE: DNA
		423	ORIGINAL COURCE:
		(VI)	ORIGINAL SOURCE:

- 75 -

			(C) INDIVIDUAL ISOLATE: AUSS	
5		AACGGC GCCATC	SEQUENCE DESCRIPTION: SEQ ID NO: 25 GCTG GTAGTAGCTC AGCTGCTCAG GGTCCCGCAA GTGG ACATGATCGC TGGTGCCCAC TGGGGAGTCC GCAT AGCGTATTTT	40 80 100
	(2)	INFORM	ATION FOR SEQ ID NO: 26	
LO		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 100 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: US4	
20		gacag gccgt	SEQUENCE DESCRIPTION: SEQ ID NO: 26 CCCTA GTGGTATCGC AGTTACTCCG GATCCCACAA CATGG ATATGGTGGC GGGGGCCCAC TGGGGAGTCC	40 80 100
25	(2)	INFOR	MATION FOR SEQ ID NO: 27	

- 76 -

		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
10			(C) INDIVIDUAL ISOLATE: ARG2	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 27	
		AGCAGC	CCTA GTGGTGTCGC AGTTACTCCG GATCCCACAA	40
		AGCATC	GTGG ACATGGTGGC GGGGGCCCAC TGGGGAGTCC	80
15			GCCT TGCTTACTAT	100
	(2)	INFORM	TATION FOR SEQ ID NO: 28	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	

- 77 -

			(C) INDIVIDUAL ISOLATE: 115	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 28	
		GGCAGC	CCTA GTGGTGTCGC AGTTACTCCG GATCCCGCAA	40
5		GCTGTC	GTGG ACATGGTGGC GGGGGCCCAC TGGGGAATCC	80
			GTCT TGCCTACTAT	100
	(2)	INFORM	NATION FOR SEQ ID NO: 29	
10		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
15				
		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: GHB	
20				
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 29	
		TGTGGG	STATE GTGGTGGCGC ACGTCCTGCG TTTGCCCCAG	40
		ACCTTG	STTCG ACATAATAGC CGGGGCCCAT TGGGGCATCT	80
		TGGCGG	GCTT GGCCTATTAC	100
25				
	(2)	INFORM	MATION FOR SEQ ID NO: 30	

•		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
10			(C) INDIVIDUAL ISOLATE: 14	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 30	
		TGTGG	STATG GTGGTAGCAC ACGTCCTGCG TCTGCCCCAG	40
		ACCTT	STICG ACATAATAGC CGGGGCCCAT TGGGGCATCT	80
15			GGCCT AGCCTATTAC	100
	(2)	INFOR	MATION FOR SEQ ID NO: 31	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 100 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	

- 79 -

	`.	(C) INDIVIDUAL ISOLATE: 111	
5		ACCITGITCG ACGIGCTAGC CGGGGCCCA1 10000011-01	40 80 100
	(2)	INFORMATION FOR SEQ ID NO: 32	
10		<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 100 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
15		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: I10	
20		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32 TACCACTATG CTCCTGGCAT ACTTGGTGCG CATCCCGGAG GTCATCCTGG ACATTATCAC GGGAGGACAC TGGGGCGTGA TGTTTGGCCT GGCTTATTTC	40 80 100
25	(2)	INFORMATION FOR SEQ ID NO: 33	

- 80 -

		(i). SEQUENCE CHARACTERISTICS:
		(A) LENGTH: 252 nucleotides
		(B) TYPE: nucleic acid
		(C) STRANDEDNESS: single
5		(D) TOPOLOGY: linear
		(ii) MOLECULE TYPE: DNA
		(vi) ORIGINAL SOURCE: (ATCC # 40394)
10		(C) INDIVIDUAL ISOLATE: hcvl
15		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33  GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC  CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACCC  GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC  GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG  CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTCTGGTAC  TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT  24  AGACCGTGCA CC
20	(2)	INFORMATION FOR SEQ ID NO: 34
		<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 252 nucleotides</li><li>(B) TYPE: nucleic acid</li></ul>
25		(C) STRANDEDNESS: single
		(D) TOPOLOGY: linear

1

- 81 <del>-</del>

	(ii) MOLECULE TYPE: DNA	
	(vi) ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: us5	
10	(XI) SEQUENCE DESCRIPTION: SEQ ID NO: 34 GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT AGACCGTGCA CC	40 80 120 160 200 240 252
15 (2)	INFORMATION FOR SEQ ID NO: 35	
20	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 252 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
25	(vi) ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: ausl	

- 82 -

		(xi) SEQUENCE DESCRIPTION. SEQ 15 No. 00	
		GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
_		GCTCAATGCC TGGAGATTTG GGCACGCCCC CGCAAGATCA	160
5		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGTGCA CC	252
		AGACCGIGCA CC	
10	(2)	INFORMATION FOR SEQ ID NO: 36	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
15		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: 5P2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36	
		GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC	40
25		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATAAACCC	120
		GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160

\_ 92 \_

		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGTGCA CC	252
5	(2)	INFORMATION FOR SEQ ID NO: 37	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
15		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: gm2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37	
		GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
20		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
		GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG	160
		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
25		AGACCGTGCA CC	252
	(2)	INFORMATION FOR SEC ID NO. 38	

- 84 -

		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
10			(C) INDIVIDUAL ISOLATE: 121	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 38	
			ATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	40
		CGGGAG	AGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
15		GGAATT	GCCA GGACGACCGG GTCCTTTCTT GGATAAACCC	120
			TGCC TGGAGATTTG GGCGTGCCCC CGCAAGACTG	160
		CTAGCC	GAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTG	ATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCG	TGCA CC	252
20				
	(2)	INFORM	ATION FOR SEQ ID NO: 39	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
25			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	

- 85 -

		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
_			(C) INDIVIDUAL ISOLATE: us4	
5				
			SEQUENCE DESCRIPTION: SEQ ID NO: 39	
•	• .	•	ATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC	
			AGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	
		GGAATT	GCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
10		GCTCAAT	IGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160
		CTAGCC	GAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
	•	TGCCTG	ATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
_		AGACCG	IGCA CC	252
15	(2)	INFORM	ATION FOR SEQ ID NO: 40	
	,	(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
20			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
25		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: jhl	

- 86 -

		(xi) SEQUENCE DESCRIPTION: SEQ 1D NO. 15	
	•	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC	40
		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
		GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160
5		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
			252
		AGACCGTGCA TC	
	4-3	INFORMATION FOR SEQ ID NO: 41	
10	(2)	INFORMATION FOR BEG ID 110	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 252 nucleotides (B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
15		(D) TOPOLOGY: linear	
		(D) TOPOLOGI. Illiedi	
		MARKET THE TOTAL	
		(ii) MOLECULE TYPE: DNA	
		and a course of the course of	
20		<pre>(vi) ORIGINAL SOURCE:     (C) INDIVIDUAL ISOLATE: nac5</pre>	
		(C) INDIVIDUAL ISOLATE: Hads	
		TROUGHT TO TO MO! AT	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41	40
		GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC	80
25		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	120
	•	GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	
		GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160

- 87 -

		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGTGCA CC	<b>2</b> 52
5	(2)	INFORMATION FOR SEQ ID NO: 42	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
10		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	
15		(C) INDIVIDUAL ISOLATE: arg2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42	
		GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC	40
		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
20		GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
		GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160
		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGTGCA CC	252
25			
	(2)	INFORMATION FOR SEQ ID NO: 43	

- 88 -

		(i)	SEQUENCE CHARACTERISTICS:	
		(-,	(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
10			(C) INDIVIDUAL ISOLATE: spl	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 43	
		GTTAGT	ATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC	40
		CGGGAG	AGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
15		CCAATT	GCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	120
12		CCTCAA	TGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	160
		CTAGCC	GAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTG	ATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
			TGCA CC	252
20		••••	•	
20	(2)	INFORM	ATION FOR SEQ ID NO: 44	
		(i)	SEQUENCE CHARACTERISTICS:	
		,	(A) LENGTH: 252 nucleotides	
25			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	

.

- 89 -

		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: ghl	
5				
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:	44
		GTTAGT	ATGA GTGTCGTGCA GCCTCCAGGA CCCCCC	CTCC 40
		CGGGAG	AGCC ATAGTGGTCT GCGGAACCGG TGAGTA	CACC 80
		GGAATT	GCCA GGACGACCGG GTCCTTTCTT GGATCA	ACCC 120
10		GCTCAA	TGCC TGGAGATTTG GGCGTGCCCC CGCGAG	ACTG 160
		CTAGCC	GAGT AGTGTTGGGT CGCGAAAGGC CTTGTG	GTAC 200
		TGCCTG	ATAG GGTGCTTGCG AGTGCCCCGG GAGGTC	TCGT 240
		AGACCG	TGCA CC	252
15	(2)	INFORM	ATION FOR SEQ ID NO: 45	·
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
20			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
25		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: 115	•

- 90 -

		(xi)	SEQU	ENCE	DESCR:	[PTION	: SEQ	ID NO:	45	
		GTTAGT								40
		CGGGAG	AGCC	ATAG:	rggtct	GCGGA	ACCGG	TGAGTA	CACC	80
		GGAATT(	2003	CGAC	23 CCGG	GTCCT	TTCTT	GGATCA	ACCC	120
		GCTCAA	AJJE	GGAC		CCCCT		CCCCAG	ACTG	160
5		GCTCAA	rgcc	TGGA	SATTIG	GGCGI	3366		ramb C	200
		CTAGCC	GAGT	AGTG	ITGGGT	CGCGA	AAGGC	CTIGIC	GIAC	
	-	TGCCTG	ATAG	GGTG	CTTGCG	AGTGC	CCCGG	GAGGTC	CTCGT.	240
		AGACCG	rgca	CC						252
10	(2)	INFORM	atioi	I FOR	SEQ I	D NO:	46			
		(i)	SEQ	JENCE	CHARA	CTERIS	TICS:			
		•	(A)	LE	NGTH:	252 nu	cleot	ides		
			(B)	TY.	PE: nu	cleic	acid			
15			(C)		RANDED	NESS:	sing	le		
15			(D)		POLOGY					
			()							
		(ii)	MOL	ECULE	TYPE:	DNA				
20		(vi)	ORI	GINAL	SOURC	E:				
		•	(C)	IN	DIVIDU	AL ISC	LATE:	<b>i10</b>		
			/							

- 91 -

		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 46	
		GCTAGTA	TCA GTGTCGTACA GCCTCCAGGC CCCCCCCTCC	40
			GCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
			CCG GGAAGACTGG GTCCTTTCTT GGATAAACCC	120
5			GCC CGGCCATTTG GGCGTGCCCC CGCAAGACTG	160
•			AGT AGCGTTGGGT TGCGAAAGGC CTTGTGGTAC	200
			TAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
		AGACCGI		252
10	(2)	INFORMA	TION FOR SEQ ID NO: 47	
	\-,			
		(i)	SEQUENCE CHARACTERISTICS:	
		•	(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
15			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
			·	
		(ii)	MOLECULE TYPE: DNA	
20		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: arg6	
			•	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 47	
		GTTAGTA	ATGA GTCTCGTACA GCCTCCAGGC CCCCCCTCC	40
25		CGGGAGA	AGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		GGAATTO	CTG GGAAGACTGG GTCCTTTCTT GGATAAACCC	120
		ACTCTAT	GCC CAGCCATTTG GGCGTGCCCC CGCAAGACTG	160
			•	

- 92 -

		CTAGCCGAGT AGCGTTGGGT TGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	-
		AGACCGTGCA TC	252
		TOD SEO ID NO: 48	
5	(2)		
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 252 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
10		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	
15		(C) INDIVIDUAL ISOLATE: 521	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48	
		GTTAGTACGA GTGTCGTGCA GCCTCCAGGA CTCCCCTCC	40
20		CGGGAGAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
20		GGAATCGCTG GGGTGACCGG GTCCTTTCTT GGAGCAACCC	120
	•	GCTCAATACC CAGAAATTTG GGCGTGCCCC CGCGAGATCA	160
		CTAGCCGAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC	200
		TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT	240
			252
25		AGACCGTGCA AC	
	(2)	INFORMATION FOR SEO ID NO: 49	

- 93 -

		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 252 nucleotides	
			(B) TYPE: nucleic acid	
5			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
10		(vi)	ORIGINAL SOURCE:	
			(C) INDIVIDUAL ISOLATE: gj61329	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 49	
15		GTTAGT	ACGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC	4 (
		CGGGAG	BAGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	8 (
		GGAATC	GCTG GGGTGACCGG GTCCTTTCTT GGAGTAACCC 19	20
		GCTCAA	TACC CAGAAATTTG GGCGTGCCCC CGCGAGATCA 1	60
		CTAGCC	GAGT AGTGTTGGGT CGCGAAAGGC CTTGTGGTAC 2	0 (
20		TGCCTG	SATAG GGTGCTTGCG AGTGCCCCGG GAGGTCTCGT 2	4 (
		AGACCG	STGCA AC 2	52
	(2)	INFORM	MATION FOR SEQ ID NO: 50	
25		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 180 nucleotides	

- 94 -

	:		<ul><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: single</li><li>(D) TOPOLOGY: linear</li></ul>	
5			MOLECULE TYPE: DNA ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: sa3	
7.0		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 50	
10		GTTAGT	ATGA GTGTCGAACA GCCTCCAGGA CCCCCCTCC	40
		CCCCAG	AGCC ATAGTGGTCT GCGGAACCGG TGAGTACACC	80
		ссаатт	GCCG GGATGACCGG GTCCTTTCTT GGATAAACCC	120
		GCTCAA	TGCC CGGAGATTTG GGCGTGCCCC CGCGAGACTG	160
15		CTAGCC	GAGT AGTGTTGGGT	180
	(2)	INFORM	ATION FOR SEQ ID NO: 51	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 180 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(vi)	ORIGINAL SOURCE:	

- 95 -

			(C)	IN	DIVIDU	AL :	ISOL	ATE:	584		
		(xi)	SEQUE	NCE	DESCR	IPT:	ION:	SEQ	ID NO	: 51	
			TATGA G								40
5		_	BAGCC A								
		GGAATI	recce e	GAT	GACCGG	GT	CCTT'	TCTT	GGATA	AACCC	120
		GCTCA	ATGCC C	GGA	GATTTG	GG	CGTG	cccc	CGCGA	GACTG	160
		CTAGCO	GAGT A	GTG:	TTGGGT						180
10											
	(2)	INFORM	MATION	FOR	SEQ I	D N	D: 5	2			
	•	(i)	SEQUE	NCE	CHARA	CTE	RIST	ics:			
			(A)	LE	ngth:	549	nuc	leot	ides		
15			(B)	TY	PE: nu	cle	ic a	cid			
			(C)	ST	RANDED	NES	s:	sing	le		
			(D)	TO	POLOGY	: 1	inea	r			
		(ii)	MOLEC	CULE	TYPE:	D	NA				
20			•								
		(vi)	ORIGI	NAL	SOURC	E:	(AT	CC #	40394	1)	
			(C)	IN	DIVIDU	AL	ISOL	ATE:	hcvl	L	

- 96 -

		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52	
		(xi) SEQUENCE DESCRIPTION: SEQ 15 NOT COMPANY ATGRAGACGA ATCCTARACC TCARARARA ARCARACGTA	40
		ATGAGCACGA ATCCTAAACC TCAAAACT TCCCGGGTGG	80
		ACACCAACCG TCGCCCACAG GACGTCAAGT TCCCGGGTGG	120
		CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG	160
5		GGCCCTAGAT TGGGTGTGCG CGCGACGAGA AAGACTTCCG	200
		AGCGGTCGCA ACCTCGAGGT AGACGTCAGC CTATCCCCAA	240
		GGCTCGTCGG CCCGAGGGCA GGACCTGGGC TCAGCCCGGG	
		TACCUTUGGE CECTETATES CAATGAGGGE TGCGGGTGGG	280
		CGGGATGGCT CCTGTCTCCC CGTGGCTCTC GGCCTAGCTG	320
10		GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT	360
10		AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA	400
		TGGGGTACAT ACCGCTCGTC GGCGCCCCTC TTGGAGGCGC	440
		TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC	480
		GGCGTGAACT ATGCAACAGG GAACCTTCCT GGTTGCTCTT	520
		TCTCTATCTT CCTTCTGGCC CTGCTCTCT	549
15		TCTCTATCTT CCTTCTGGCC C2CCC	
	(2)	INFORMATION FOR SEQ ID NO: 53	
		THE PART OF THE PA	
		(i) SEQUENCE CHARACTERISTICS:	
20		(A) LENGTH: 549 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
25		(ii) MOLECULE TYPE: DNA	
		(vi) ORIGINAL SOURCE:	

(C) INDIVIDUAL	ISOLATE: us5
----------------	--------------

	(xi) SEQU	JENCE DESCRI	IPTION: SEQ	ID NO: 23	
	ATGAGCACGA	ATCCTAAACC	TCAAAGAAAA	ACCAAACGTA	40
5	ACACCAACCG	TCGCCCACAG	GACGTCAAGT	TCCCGGGTGG	80
	CGGTCAGATC	GTTGGTGGAG	TTTACTTGTT	GCCGCGCAGG	120
	GGCCCTAGAT	TGGGTGTGCG	CGCGACGAGG	AAGACTTCCG	160
	AGCGGTCGCA	ACCTCGAGGT	AGACGTCAGC	CTATCCCCAA	200
	GGCGCGTCGG	CCCGAGGGCA	GGACCTGGGC	TCAGCCCGGG	240
10	TACCCTTGGC	CCCTCTATGG	CAATGAGGGT	TGCGGGTGGG	280
	CGGGATGGCT	CCTGTCTCCC	CGTGGCTCTC	GGCCTAGTTG	320
	GGGCCCCACA	GACCCCGGC	GTAGGTCGCG	CAATTTGGGT	360
	AAGGTCATCG	ATACCCTTAC	GTGCGGCTTC	GCCGACCACA	400
	TGGGGTACAT	ACCGCTCGTC	GGCGCCCCTC	TTGGAGGCGC	440
15	TGCCAGGGCT	CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC	480
		ATGCAACAGG			520
		CCTTCTGGCC			549

## (2) INFORMATION FOR SEQ ID NO: 54

20

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 549 nucleotides
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
- 25 (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA

- 98 - .

		(vi)	ORIG		SOURC							
			(C)	IN	DIVIDU	AL	ISOL	ATE:	aus	1		
_		(xi)	SEOI	TENCE	DESCR	IPT	ion:	SEQ	ID N	0:	54	
5		ATGAGC	, 200 y	אשרכי	רבבב	тC	AAAG	AAAA	ACCA	AAC	GTA	40
		ACACCA	ACGA	DOCC.		G)		ልልርጥ	TCCC	:GGG	TGG	80
	•	ACACCA	ACCG	TUGU		. Un		ייים איים	GCCG	ירפר	AGG	120
		CGGTCA	BATC	GTTG	GTGGAG	TT	TACI.	1911	3363	CET	icce	160
		GGCCCT	AGAT	TGGG'	TGTGCG	CG	CGAC	GAGG	AAGA	CTI		200
10		AGCGGT	CGCA	ACCT	CGAGGT	AG	ACGT	CAGC	CTAT	CCC.	TAA	
		GGCGCG!	rcgg	CCCG	AGGGCA	GG	ACCT	GGGC	TCAG	CCC	GGG	240
		TACCCC'	rggc	CCCT	CTATGG	TA	ATGA	gggt	TGCG	GAT	GGG	280
		CGGGAT	GGCT.	CCTG	TCCCCC	CG	TGGC	TCTC	GGCC	TAG	TTG	320
		GGGCCC	מחמיי	GACC	CCCGGC	GI	AGGT	CGCG	CAAT	TTC	GGT	360
		AAGGTC	1860	3830		GT.	יפרפפ	СФФС	GCCG	ACC	ACA	400
15		AAGGTC	atcg	ATAC	CCICAC	. 61	.0000	COMO	TTGG	2000	CGC	440
		TGGGGT	ACAT	TCCG	CTCGTT	G			1100	7000	7000 7000	480
		TGCCAG	GGCC	CTGG	CGCATG	GC	GTCC	GGGT	TCTC		AGAC	
		GGCGTG	AACT	ATGC	AACAGG	G.P	ATCT	TCCT	GGTI	[GC]	CTT	520
		TCTCTA	TCTT	CCTT	CTGGCC	: CI	TCTC	TCT				549
20												
20	(2)	INFORM	ATIO	n for	SEQ I	D N	NO: 5	5				
		(i)	SEQ	UENCE	CHARA	CTI	ERIST	ics:				
			(A)	LE	ngth:	549	סטת פ	leot	ides			
25			(B)	TY	PE: nu	cle	eic a	cid				
25			(C)		RANDEI				le			
			( )					-				

- 99 -

			(D)	TO	POLOGY	: linea:	r		
		(ii)	MOLE	CULE	TYPE:	DNA			
5		(vi)							
						AL ISOL			
		(xi)	SEQU	JENCE	DESCR	IPTION:	SEQ	ID NO: 55	
		ATGAGC	ACGA	ATCC:	TAAACC	TCAAAG	AAAA	ACCAAACGTA	
		ACACCA	ACCG	TCGC	CCACAG	GACGTC	AAGT	TCCCGGGTGG	80
10		CGGTCA	GATC	GTTG	GTGGAG	TTTACT	TGTT	GCCGCGCAGG	120
		GGCCCT	AGAT	TGGG:	rgtgcg	CACGAC	GAGG	AAGACTTCCG	160
		AGCGGT	CGCA	ACCT	CGAGGT	AGACGT	CAGC	CCATCCCCAA	200
		GGCTCG	TCGA	CCCG	AGGGCA	GGACCT	GGGC	TCAGCCCGGG	240
		TACCCT	TGGC	CCCT	CTATGG	CAATGA	GGGC	TGCGGGTGGG	280
15		CGGGAT	GGCT	CCTG!	TCTCCC	CGTGGC	TCTC	GGCCTAGCTG	320
		GGGCCC	CACA	GACC	CCCGGC	GTAGGT	CGCG	CAATTTGGGT	360
		AAGGTC	ATCG	ATAC	CCTTAC	GTGCGG	CTTC	GCCGACCTCA	400
		TGGGGT.	ACAT	ACCG	CTCGTC	GGCGCC	CCTC	TTGGAGGCGC	440
		TGCCAG.	AGCC	CTGG	CGCATG	GCGTCC	GGGT	TCTGGAAGAC	480
20		GGCGTG.	AACT	ATGC	AACAGG	GAACCT'	TCCC	GGTTGCTCTT	520
		TCTCTA	TCTT	CCTT	CTGGCC	CTGCTC	TCT		549
	(2)	INFORM	OITA	N FOR	SEQ I	D NO: 5	6		
25		(i)	SEQ	JENCE	CHARA	CTERIST	ics:		
			(A)	LE	ngth :	549 nuc	leot	ides	
			(B)	TY	PE: nu	cleic a	cid		

- 100 -

		(C) STRANDEDNESS: single
		(D) TOPOLOGY: linear
-		(ii) MOLECULE TYPE: DNA
5		(vi) ORIGINAL SOURCE: (C) INDIVIDUAL ISOLATE: gm2
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56
10		ATGAGCACGA ATCCTAAACC TCAAAGARGA ACCTISTO
		ACACCAACCG TCGCCCACAG GACGICAAGI ICCCCCTTC
		CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG 120
		GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG 160
		AGCGGTCGCA ACCTCGAGGT AGACGTCAGC CTATCCCCAA 200
15		GGCACGTCGG CCCGAGGGTA GGACCTGGGC TCAGCCCGGG 240
		TACCCTTGGC CCCTCTATGG CAATGAGGGT TGCGGGTGGG 280
		CGGGATGGCT CCTGTCTCCC CGCGGCTCTC GGCCTAACTG 320
		GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT 360
		AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA 400
		TGGGGTACAT ACCGCTCGTC GGCGCCCCTC TTGGAGGCGC 440
20		TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC 480
		GGCGTGAACT ATGCAACAGG GAACCTTCCT GGTTGCTCTT 520
		TCTCTATCTT CCTTCTGGCC CTGCTCTCT 549
		TCTCTATCTT CCTTCTGGCC CTGGTGTG
25	(2)	INFORMATION FOR SEQ ID NO: 57
		(i) SEQUENCE CHARACTERISTICS:

- 101 -

		(A) LENGTH: 549 nucleotides
		(B) TYPE: nucleic acid
		(C) STRANDEDNESS: single
		(D) TOPOLOGY: linear
5		
		(ii) MOLECULE TYPE: DNA
		(vi) ORIGINAL SOURCE:
		(C) INDIVIDUAL ISOLATE: 121
10		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57
		ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA 4
		ACACCAACCG TCGCCCACAG GACGTCAAGT TCCCGGGTGG 8
		CGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG 12
		GGCCCTAGAT TGGGTGTGCG CGCGACGAGG AAGACTTCCG 16
15		AGCGGTCGCA ACCTCGTGGT AGACGCCAGC CTATCCCCAA 20
		GGCGCGTCGG CCCGAGGGCA GGACCTGGGC TCAGCCCGGG 24
		TACCCTTGGC CCCTCTATGG CAATGAGGGT TGCGGGTGGG 28
		CGGGATGGCT CCTGTCTCCC CGTGGCTCTC GGCCTAGCTG 32
		GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT 36
20		AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA 40
		TGGGGTACAT ACCGCTCGTC GGCGCCCCTC TTGGAGGCGC 44
		TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC 48
		GGCGTGAACT ATGCAACAGG GAACCTTCCT GGTTGCTCTT 52
		TTTCTATTTT CCTTCTGGCC CTGCTCTCT 54
25		
	(2)	INFORMATION FOR SEQ ID NO: 58

- 102 -

	(i) SEQUE	NCE CHARACTERISTICS:	
	(A)	LENGTH: 549 nucleotides	
	(B)	TYPE: nucleic acid	
5	(C)	STRANDEDNESS: single	
	(D)	TOPOLOGY: linear	
	(ii) MOLEC	ULE TYPE: DNA	
	(vi) ORIGI	NAL SOURCE:	
10	(C)	INDIVIDUAL ISOLATE: us4	
	(xi) SEQUE	NCE DESCRIPTION: SEQ ID NO: 58	
•	ATGAGCACGA A	TCCTAAACC TCAAAGAAAA ACCAAACGTA	40
	ACACCAACCG C	CGCCCACAG GACGTTAAGT TCCCGGGCGG	80
15	TGGCCAGGTC G	TTGGTGGAG TTTACCTGTT GCCGCGCAGG	120
	GGCCCCAGGT T	GGGTGTGCG CGCGACTAGG AAGACTTCCG	160
	AGCGGTCGCA A	ACCTCGTGGA AGGCGACAAC CTATCCCCAA	200
	GGCTCGCCAG C	CCGAGGGCA GGGCCTGGGC TCAGCCCGGG	240
	TACCCTTGGC C	CCTCTATGG CAATGAGGGT ATGGGGTGGG	280
20	CAGGATGGCT C	CCTGTCACCC CGTGGCTCTC GGCCTAGTTG	320
20	GGGCCCCACG G	SACCCCCGGC GTAGGTCGCG TAATTTGGGT	360
	AAGGTCATCG A	ATACCCTCAC ATGCGGCTTC GCCGACCTCA	400
	TGGGGTACAT I	CCGCTCGTC GGCGCCCCCC TTAGGGGCGC	440
	TGCCAGGGCC I	TTGGCGCATG GCGTCCGGGT TCTGGAGGAC	48
25	GGCGTGAACT A	ACGCAACAGG GAATCTGCCC GGTTGCTCCT	52
-•		CCTCTTGGCT CTGCTGTCC	549

200

240

280 320

360

400

440

480 520

5

10

15

20

25

- 103 -

(2) INFORMATION FOR SEQ ID NO: 59

(i) !	EQUENCE (	HARAC	TERISTIC	s:		
(	A) LENG	TH: 5	49 nucle	otides	5	
(	B) TYPE	e: nuc	leic aci	đ		
(	C) STR	MDEDN	ESS: si	ngle		
. (	D) TOPO	)LOGY:	linear			
(ii) 1	OLECULE 1	YPE:	DNA			
(vi) (	RIGINAL 8	OURCE	<b>:</b> :			
(	C) INDI	VIDUA	L ISOLAT	E: jl	ıl.	
(xi) 8	EQUENCE I	ESCRI	PTION: S	EQ ID	NO: 59	)
ATGAGCA(	AA ATCCTA	LAACC	TCAAAGAA	AA ACC	CAAACGI	A 40
ACACCAA	ce ccecco	ACAG	GACGTCAA	GT TCC	CGGGCG	G 80
TGGTCAG	TC GTTGGT	:GGAG	TTTACCTG	TT GCC	CGCGCAG	G 120
GGCCCCA	GT TGGGT	TGCG	CGCGACTA	GG AAG	SACTTCC	G 160

AGCGGTCGCA ACCTCGTGGA AGGCGACAAC CTATCCCCAA GGCTCGCCAG CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG

TACCCTTGGC CCCTCTATGG CAACGAGGGT ATGGGGTGGG

CAGGATGGCT CCTGTCACCC CGTGGCTCTC GGCCTAGTTG

GGGCCCCACG GACCCCCGGC GTAGGTCGCG TAATTTGGGT

AAGGTCATCG ATACCCTCAC ATGCGGCTTC GCCGACCTCA
TGGGGTACAT TCCGCTTGTC GGCGCCCCCC TAGGGGGCGC

TGCCAGGGCC CTGGCACATG GTGTCCGGGT TCTGGAGGAC

GGCGTGAACT ATGCAACAGG GAATTTGCCC GGTTGCTCTT

- 104 -

		TCTCTATCT	CCTCTTGGCT CTGCTGTCC	549
	(2)	INFORMATIO	ON FOR SEQ ID NO: 60	
5		(A) (B) (C)	QUENCE CHARACTERISTICS: LENGTH: 549 nucleotides TYPE: nucleic acid STRANDEDNESS: single	
10			TOPOLOGY: linear LECULE TYPE: DNA	
		(12)	GINAL SOURCE: INDIVIDUAL ISOLATE: nac5	
15		(xi) SEQ	QUENCE DESCRIPTION: SEQ ID NO: 60	
			A ATCCTABACC CCABAGABAB ACCABACGTA	40
			TCGCCCACAG GACGTCAAGT TCCCGGGCGG	80
			GTTGGTGGAG TTTACCTGTT GCCGCGCAGG	120
			TGGGTGTGCG CGCGACTAGG AAGACTTCCG	160
20			A ACCTCGTGGA AGGCGACAAC CTATCCCCAA	200
			CCCGAGGGCA GGTCCTGGGC TCAGCCCGGG	240
			CCCTCTATGG CAACGAGGGT ATGGGGTGGG	280
			CCTGTCACCC CGCGGCTCCC GGCCTAGTTG	320
			GACCCCGGC GTAGGTCGCG TAATTTGGGT	360
25				400
		AAGGTCATC	ATACCCTCAC ATGCGGCTTC GCCGACCTCA	400

- 105 -

		TGGGGTACAT TCCGCTCGTC GGCGCCCCCC TAGGGGGCGC	440
		TGCCAGGGCC CTGGCACATG GTGTCCGGGT TCTGGAGGAC	480
		GGCGTGAACT ATGCAACAGG GAATTTGCCT GGTTGCTCTT	520
		TCTCTATCTT CCTCTTGGCT CTGCTGTCC	549
5		•	
	(2)	INFORMATION FOR SEQ ID NO: 61	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 549 nucleotides	
		(B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
15		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: arg2	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61	
		ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA	40
20		ACACCAACCG CCGCCCACAG GACGTCAAGT TCCCGGGCGG	80
		TGGTCAGATC GTTGGTGGAG TTTACTTGTT GCCGCGCAGG	120
		GGCCCCAGGT TGGGTGTGCG CGCGACTAGG AAGACTTCCG	160
		AGCGGTCGCA ACCTCGTGGA AGGCGACAAC CTATCCCCAA	200
		GGCTCGCCAG CCCGAGGGTA GGGCCTGGGC TCAGCCCGGG	240
25		TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG	280
		CAGGGTGGCT CCTGTCCCCC CGCGGCTCCC GGCCTAGTTG	320

- 108 -

		GGCTCGCCGG CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG	240
		TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG	280
		CAGGATGGCT CCTGTCACCC CGTGGTTCTC GGCCTAGTTG	320
		GGGCCCCACG GACCCCCGGC GTAGGTCGCG CAATTTGGGT	360
_		AAGATCATCG ATACCCTCAC GTGCGGCTTC GCCGACCTCA	400
5		TGGGGTACAT TCCGCTCGTC GGCGCCCCCC TAGGGGGCGC	440
		TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAGGAC	480
		GGCGTGAACT ATGCAACAGG GAATCTGCCC GGTTGCTCCT	520
			549
		TTTCTATCTT CCTTCTGGCT TTGCTGTCC	
10			
	(2)	INFORMATION FOR SEQ ID NO: 64	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 549 nucleotides	
15		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20			
		(vi) ORIGINAL SOURCE:	
		(C) INDIVIDUAL ISOLATE: i15	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64	
25		ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA	40
23		ACACCAACCG CCGCCCACAG GACGTCAAGT TCCCGGGCGG	80
		TGGTCAGATC GTTGGTGGAG TTTACCTGTT GCCGCGCAGG	120

**- 109 -**

		GGCCC	CAGGT	TGGGTGTGCG	CGCGACTAGG	AAGACTTCCG	160
		AGCGG	TCGCA	ACCTCGTGGA	AGGCGACAAC	CTATCCCCAA	200
		GGCTC	GCCAG	CCCGAGGGCA	GGGCCTGGGC	TCAGCCCGGG	240
		TACCC	CTGGC	CCCTCTATGG	CAATGAGGGT	ATGGGGTGGG	280
5		CAGGA	TGGCT	CCTGTCACCC	CGCGGCTCCC	GGCCTAGTTG	320
•						TAATTTGGGT	
		AAGGT	CATCG	ATACCCTCAC	ATGCGGCTTC	GCCGACCTCA	400
		TGGGG	TACAT	TCCGCTCGTC	GGCGCCCCT	TAGGGGGCGC	440
		TGCCA	GGGCC	CTGGCGCATG	GCGTCCGGGT	TCTGGAGGAC	480
10		GGCGT	GAACT	ATGCAACAGG	GAATCTACCC	GGTTGCTCTT	520
		TCTCT	ATCTT	CCTCTTGGCT	TTGCTGTCC		549
	(2)	INFOR	MATIO	N FOR SEQ I	D NO: 65		
15		(i)	SEQU	JENCE CHARA	CTERISTICS:		
			(A)	LENGTH:	549 nucleot	ides	
			(B)	TYPE: nuc	cleic acid		
			(C)	STRANDEDI	NESS: sing	le	
			(D)	TOPOLOGY	: linear		
20							
		(ii)	MOLE	CULE TYPE:	DNA		
		(vi)	ORIG	SINAL SOURCE	∃:		
		( /	(C)		AL ISOLATE:	110	
25			, -,		·		
		(xi)	SEQU	ENCE DESCR	PTION: SEQ	ID NO: 65	
		ATGAG	CACAA	ATCCTAAACC	TCAAAGAAAA	ACCAAAAGAA	40

- 110 -

						mecceaces	80
		ACACTAA	CCG C	CGCCCACAG	GACGTCAAGT	100000000	120
		TGGCCAG	ATC G	TTGGCGGAG	TATACTTGCT	GCCGCGCAGG	
		GGCCCGA	GAT T	GGGTGTGCG	CGCGACGAGG	AAAACTTCCG	160
		AACGATC	CCA G	CCACGCGGA	AGGCGTCAGC	CCATCCCTAA	200
5		AGATCGT	CGC A	CCGCTGGCA	AGTCCTGGGG	AAGGCCAGGA	240
J		татсстт	GGC C	CCTGTATGG	GAATGAGGGT	CTCGGCTGGG	280
		CAGGGTG	GCT C	CTGTCCCCC	CGTGGCTCTC	GCCCTTCATG	320
		GGGCCCC	ACT G	ACCCCCGGC	ATAGATCGCG	CAACTTGGGT	360
		AAGGTCA	TCG A	TACCCTAAC	GTGCGGTTTT	GCCGACCTCA	400
		magagata.	САТ Т	CCCGTCATC	GGCGCCCCG	TTGGAGGCGT	440
10		TGGGGT	CCT C	TCGCCCACG	GAGTGAGGGT	TCTGGAGGAT	480
		CCCCTAA	מידית ב	TGCAACAGG	GAATTTGCCC	GGTTGCTCTT	520
					CTCTTGTCT		549
		TCTCIAI	C11 1				
15	(2)	INFORMA	TION	FOR SEQ I	D NO: 66		
		(i)	SEQUE		CTERISTICS:		
			(A)		510 nucleot	ides	
			(B)		cleic acid		
20			(C)	STRANDEL	NESS: sing	le	
			(D)	TOPOLOGY	: linear		
		(ii)	MOLE	CULE TYPE:	DNA		
25		(vi)	ORIG	INAL SOURC	Œ:		
23			(C)	INDIVIDU	JAL ISOLATE:	arg6	

	(2)	INFORMATION FOR SEQ ID NO: 68	
5		<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 24 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
10		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68 ACAGAYCCGC AKAGRTCCCC CACG	24
15	(2)	INFORMATION FOR SEQ ID NO: 69	
20		<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 30 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
		(ii) MOLECULE TYPE: DNA	
25		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69	30

	(2)	INFOR	MATION FOR SEQ ID NO: 70	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 30 nucleotides	
5			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
•		(ii)	MOLECULE TYPE: DNA	
10		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 70	
		GCAAC	CTCGT GGAAGGCGAC AACCTATCCC	3(
•	(2)	INFOR	MATION FOR SEQ ID NO: 71	
15		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 30 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
20			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 71	
25		GTCAC	CAATG ATTGCCCTAA CTCGAGTATT	3(
	(0)	TATEOD1	ANTON FOR SEC ID NO. 72	

		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 26 nucleotides	
			(B) TYPE: nucleic acid	
5			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
10	•	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 72	
		GTCACG	BAACG ACTGCTCCAA CTCAAG	26
	(2)	INFORM	ATION FOR SEQ ID NO: 73	
15		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 28 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
•			(D) TOPOLOGY: linear	
20				
		(ii)	MOLECULE TYPE: DNA	
			THE PERCENTAGE SEC ID NO: 73	
			SEQUENCE DESCRIPTION: SEQ ID NO: 73 ATGAT CGCTGGWGCY CACTGGGG	28
		TGGACA	AIGHI CGCIGGHOCI WASSACE	
25	(2)	THEODI	MATTON FOR SEO ID NO: 74	

		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 28 nucleotides	
		(B) TYPE: nucleic acid	
5		(C) STRANDEDNESS: single	
•		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74	
10		TGGAYATGGT GGYGGGGGCY CACTGGGG	28
	(2)	INFORMATION FOR SEQ ID NO: 75	
		(i) SEQUENCE CHARACTERISTICS:	
15		(A) LENGTH: 20 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
20		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75	
		ATGATGAACT GGTCVCCYAC	20
25	(2)	INFORMATION FOR SEQ ID NO: 76	
		(i) SEQUENCE CHARACTERISTICS:	

- 116 -

		<ul> <li>(A) LENGTH: 26 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
5		(ii) MOLECULE TYPE: DNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76 ACCTTVGCCC AGTTSCCCRC CATGGA	26
10	(2)	INFORMATION FOR SEQ ID NO: 77	
15		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 22 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77  AACCCACTCT ATGYCCGGYC AT	22
	(2)	INFORMATION FOR SEQ ID NO: 78	
25		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 18 nucleotides  (B) TYPE: nucleic acid	

- 117 -

			(C) STRANDEDNESS: single	
•			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
5				
		•	SEQUENCE DESCRIPTION: SEQ ID NO: 78	
		GAATCG	CTGG GGTGACCG	18
	(0)	***********	1810V BOD 650 15 Vo. 50	
10	(2)	INFORM	ATION FOR SEQ ID NO: 79	
10		(1)	SEQUENCE CHARACTERISTICS:	
		\-/	(A) LENGTH: 28 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
15			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
			SEQUENCE DESCRIPTION: SEQ ID NO: 75	
20		CCATGA	ATCA CTCCCCTGTG AGGAACTA	28
	(0)	7.77071	ARTON TOD STO ID NO. 00	
	(2)	INFORM	ATION FOR SEQ ID NO: 80	
		(i)	SEQUENCE CHARACTERISTICS:	
25		(-/	(A) LENGTH: 18 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	

- 118 -

			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
5	(2)	TTGCGG	SEQUENCE DESCRIPTION: SEQ ID NO: 80  GGGC ACGCCCAA  ATION FOR SEQ ID NO: 81	18
10		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
20	(2)	YGAAGC	SEQUENCE DESCRIPTION: SEQ ID NO: 81 GGGC ACAGTCARRC AAGARAGCAG GGC ATION FOR SEQ ID NO: 82	33
25		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single	

		:	(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
5		•	SEQUENCE DESCRIPTION: SEQ ID NO: 82	33
•	(2)		MATION FOR SEQ ID NO: 83	
10		• -	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
			SEQUENCE DESCRIPTION: SEQ ID NO: 83	33
20	(2)	INFORM	MATION FOR SEQ ID NO: 84	
25		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
			•	

- 120 -

		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84 AGYRTGCAGG ATGGYATCRK BCGYCTCGTA CAC	33
5	(2)	INFORMATION FOR SEQ ID NO: 85  (i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides	
10		(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
15		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85 GTTRCCCTCR CGAACGCAAG GGACRCACCC CGG	33
	(2)	INFORMATION FOR SEQ ID NO: 86	
20		<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 33 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
25		(ii) MOLECULE TYPE: DNA	

	•		SEQUENCE DESCRIPTION: SEQ ID NO: 86 GGTY AYCGCCACCC AACACCTCGA GRC	33
5	(2)	INFORM	MATION FOR SEQ ID NO: 87	
		(i)	SEQUENCE CHARACTERISTICS:	
		•	(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
		•	(C) STRANDEDNESS: single	
10			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 87	
15		CGTYG	GGGG AGTTTGCCRT CCCTGGTGGC YAC	33
	(2)	INFORM	MATION FOR SEQ ID NO: 88	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 88	

- 122 -

		CCCGACAAGC AGATCGATGT GACGTCGAAG CTG	33
	(2)	INFORMATION FOR SEQ ID NO: 89	
5	·	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single	
10		(D) TOPOLOGY: linear  (ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89 CCCCACGTAG ARGGCCGARC AGAGRGTGGC GCY	33
15	(2)	INFORMATION FOR SEQ ID NO: 90	
20		<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 33 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
25		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90	33

	(2)	INFORMATION FOR SEQ ID NO: 91	
		(i) SEQUENCE CHARACTERISTICS:	
5		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
10		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91	
		CGTCCAGTGG YGCCTGGGAG AGAAGGTGAA CAG	33
15	(2)	INFORMATION FOR SEQ ID NO: 92	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
20		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
25		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92 GCCGGGATAG ATRGARCAAT TGCARYCTTG CGT	33

- 124 -

	(2)	INFORMATION FOR SEQ ID NO: 93	
5		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
10		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93 CATATCCCAT GCCATGCGGT GACCCGTTAY ATG	33
	(2)	INFORMATION FOR SEQ ID NO: 94	
15		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid (C) STRANDEDNESS: single	
20		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
25		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94 YACCAAYGCC GTCGTAGGGG ACCARTTCAT CAT	33
	(2)	INFORMATION FOR SEQ ID NO: 95	

		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
5		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95	
10		GATGGCTTGT GGGATCCGGA GYASCTGAGC YAY	33
	(2)	INFORMATION FOR SEQ ID NO: 96	
		(i) SEQUENCE CHARACTERISTICS:	
15		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
20		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96	
		GACTCCCCAG TGRGCWCCAG CGATCATRTC CAW	33
25	(2)	INFORMATION FOR SEQ ID NO: 97	
		(4) SPONENCE CHARACTERISTICS:	

- 126 - .

		<ul><li>(A) LENGTH: 33 nucleotides</li><li>(B) TYPE: nucleic acid</li><li>(C) STRANDEDNESS: single</li><li>(D) TOPOLOGY: linear</li></ul>	
5		(ii) MOLECULE TYPE: DNA  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97  CCCCACCATG GAGAAATACG CTATGCCCGC YAG	33
10	(2)	INFORMATION FOR SEQ ID NO: 98	
15		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
20		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98 TAGYAGCAGY ACTACYARGA CCTTCGCCCA GTT	33
	(2)	INFORMATION FOR SEQ ID NO: 99	
25		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid	

			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
5		(ii)	MOLECULE TYPE: DNA	
J			SEQUENCE DESCRIPTION: SEQ ID NO: 99 GTGR GTKTCYGCGT CRACGCCGGC RAA	33
10	(2)	INFORM	ATION FOR SEQ ID NO: 100	
•		(i)	SEQUENCE CHARACTERISTICS:	
		•	(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
15			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 100	
20			TGGG ATGGTYARRC ARGASAGCAR AGC	33
	(2)	INFORM	MATION FOR SEQ ID NO: 101	
		(i)	SEQUENCE CHARACTERISTICS:	
25			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	

- 128 -

		(D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: DNA	
(2)	GTAYAY	YCCG GACRCGTTGC GCACTTCRTA AGC	33
	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: DNA	
(2)	AATRCI	TTGMG TTGGAGCART CGTTYGTGAC ATG	33
	(1)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
		(xi) GTAYAY (2) INFORM (ii) (ii) (xi) AATRCI	(ii) MOLECULE TYPE: DNA  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101 GTAYAYYCCG GACRCGTTGC GCACTTCRTA AGC  (2) INFORMATION FOR SEQ ID NO: 102  (i) SEQUENCE CHARACTERISTICS:

- 129 -

		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103 RGYRTGCATG ATCAYGTCCG YYGCCTCATA CAC	33
5			
	(2)	INFORMATION FOR SEQ ID NO: 104	
		(1) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 33 nucleotides	
	•	(B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
15		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104	
		RTTGTYYTCC CGRACGCARG GCACGCACCC RGG	33
•	(2)	INFORMATION FOR SEQ ID NO: 105	
20		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
25			
		(44) MOLECULE TYPE: DNA	

**- 130 -**

		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105 CGTGGGRGTS AGCGCYACCC AGCARCGGGA GSW 33	3
	(2)	INFORMATION FOR SEQ ID NO: 106	
5	•	(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
15		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106 YGTRGTGGGG AYGCTGKHRT TCCTGGCCGC VAR 3:	3
	(2)	INFORMATION FOR SEQ ID NO: 107	
20		<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 33 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
25		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107	

.′ຸ:

- 131 -

		CCCRACGAGC AARTCGACRT GRCGTCGTAW TGT	33
	(2)	INFORMATION FOR SEQ ID NO: 108	
5		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
•		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
10			
		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108	
		YCCCACGTAC ATAGCSGAMS AGARRGYAGC CGY	33
15	(2)	INFORMATION FOR SEQ ID NO: 109	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 33 nucleotides	
20		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
25		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109	
		CTGGGAGAYR AGRAAAACAG ATCCGCARAG RTC	33

	(2)	INFORMATION FOR SEQ ID NO: 110	
5		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
10		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110 YGTCTCRTGC CGGCCAGSBG AGAAGGTGAA YAG	33
15	(2)	INFORMATION FOR SEQ ID NO: 111	
20		<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 33 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
		(ii) MOLECULE TYPE: DNA	
25		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111 GCCGGGATAG AKKGAGCART TGCAKTCCTG YAC	33

	(2)	INFORMATION FOR SEQ ID NO: 112	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 33 nucleotides	
5		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
10		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112 CATATCCCAA GCCATRCGRT GGCCTGAYAC CTG	33
		CATATCCCA OCCATAGONI COCCATAGONI	
	(2)	INFORMATION FOR SEQ ID NO: 113	
15		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
20		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113	
25		CACTARGGCT GYYGTRGGYG ACCAGTTCAT CAT	33
	(2)	INFORMATION FOR SEO ID NO: 114	

- 134 -

_		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single	
5			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
10	٠	(xi) GACRGC	SEQUENCE DESCRIPTION: SEQ ID NO: 114 TTGT GGGATCCGGA GTAACTGCGA YAC	33
	(2)	INFORM	NATION FOR SEQ ID NO: 115	
15		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
20		(ii)	MOLECULE TYPE: DNA	
		(xi) GACTC	SEQUENCE DESCRIPTION: SEQ ID NO: 115 CCCAG TGRGCCCCCG CCACCATRTC CAT	33
25	(2)	INFOR	MATION FOR SEQ ID NO: 116	
		(i)	SEQUENCE CHARACTERISTICS:	

ſ

**- 135 -** .

			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
		•	(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
5				
_		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 116	
		SCCCAC	CATG GAWWAGTAGG CAAGGCCCGC YAG	33
10	(2)	INFORM	NATION FOR SEQ ID NO: 117	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
15			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
20		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 117	
		GAGTA	SCATC ACAATCAADA CCTTAGCCCA GTT	33
	(2)	INFORM	MATION FOR SEQ ID NO: 118	
25		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	

- 136 -

			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
_		(ii)	MOLECULE TYPE: DNA	
5		(xi) YGWCRY	SEQUENCE DESCRIPTION: SEQ ID NO: 118 EGYRG GTRTKCCCGT CAACGCCGGC AAA	33
10	(2)	INFORM	NATION FOR SEQ ID NO: 119	
10		(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
15			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 119	33
20		TCCTC	ACAGG GGAGTGATTC ATGGTGGAGT GTC	33
	(2)	INFOR	MATION FOR SEQ ID NO: 120	
		(i)	SEQUENCE CHARACTERISTICS:	
25			(A) LENGTH: 33 nucleotides	
-			(B) TYPE: nucleic acid	•
			(C) STRANDEDNESS: single	

		(D) TOPOLOGY: linear
		(ii) MOLECULE TYPE: DNA
5		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120 ATGGCTAGAC GCTTTCTGCG TGAAGACAGT AGT 33 INFORMATION FOR SEQ ID NO: 121
: 10		<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 33 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>
15		(ii) MOLECULE TYPE: DNA  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121  GCCTGGAGGC TGCACGRCAC TCATACTAAC GCC 33
20	(2)	INFORMATION FOR SEQ ID NO: 122
25		<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 33 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>

PCT/US92/04036 WO 92/19743

- 138 -

		(ii) M	OLECULE TYPE: DNA	
		(xi) S	EQUENCE DESCRIPTION: SEQ ID NO: 122 AC TATGGCTCTY CCGGGAGGGG GGG	33
5	(2)	(i) S	PION FOR SEQ ID NO: 123 SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 nucleotides	
10			(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
			MOLECULE TYPE: DNA	
15		(xi) TCRTCCY	SEQUENCE DESCRIPTION: SEQ ID NO: 123 GGC AATTCCGGTG TACTCACCGG TTC	33
	(2)	INFORM	ATION FOR SEQ ID NO: 124	
20		(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	

- 139 -

			SEQUENCE DESCRIPTION: SEQ ID NO: 124 AGCG GGTTDATCCA AGAAAGGACC CGG	33
	(2)	INFORM	ATION FOR SEQ ID NO: 125	
5		(i)	SEQUENCE CHARACTERISTICS:	
		<b>\-</b> ,	(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
10			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 125	
15		AGCAGT	CTYG CGGGGGCACG CCCAARTCTC CAG	33
	(2)	INFORM	ATION FOR SEQ ID NO: 126	
		(i)	SEQUENCE CHARACTERISTICS:	
20			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
25		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 126	

WO 92/19743 PCT/US92/04036

- 140 -

		ACAAGGCCTT TCGCGACCCA ACACTACTCG GCT	33
	(2)	INFORMATION FOR SEQ ID NO: 127	
5		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
10		(ii) MOLECULE TYPE: DNA	
		(XI) SEQUENCE DESCRIPTION: SEQ ID NO: 127 GGGGCACTCG CAAGCACCCT ATCAGGCAGT ACC	33
15	(2)	INFORMATION FOR SEQ ID NO: 128	
20		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	

		(ii) MOLECULE TYPE: DNA	
5		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128 YGTGCTCATG RTGCACGGTC TACGAGACCT CCC	33
	(2)	INFORMATION FOR SEQ ID NO: 129	
10		<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 33 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
15		(ii) MOLECULE TYPE: DNA	
	•	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129 GTTACGTTTG KTTYTTYTTT GRGGTTTRGG AWT	33
20	(2)	INFORMATION FOR SEQ ID NO: 130	
		<ul><li>(i) SEQUENCE CHARACTERISTICS:</li><li>(A) LENGTH: 33 nucleotides</li><li>(B) TYPE: nucleic acid</li></ul>	
25		<pre>(C) STRANDEDNESS: single (D) TOPOLOGY: linear</pre>	

-142 -

		(ii) MOLECULE TYPE: DNA	
	·	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130 CGGGAACTTR ACGTCCTGTG GGCGRCGGTT GGT	33
5	(2)	INFORMATION FOR SEQ ID NO: 131	
10		(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
15		(ii) MOLECULE TYPE: DNA  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131  CARGTAAACT CCACCRACGA TCTGRCCRCC RCC	33
20	(2)	INFORMATION FOR SEQ ID NO: 132  (i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 33 nucleotides  (B) TYPE: nucleic acid	
25		(C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA	

- 143 -

		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132 RCGCACACCC AAYCTRGGGC CCCTGCGCGG CAA	33
5	(2)	INFORMATION FOR SEQ ID NO: 133	
		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
10		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133	
15		AGGTTGCGAC CGCTCGGAAG TCTTYCTRGT CGC	33
	(2)	INFORMATION FOR SEQ ID NO: 134	
		(i) SEQUENCE CHARACTERISTICS:	
20		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
25		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134	

- 144 -

		RCGHR	CCTTG GGGATAGGCT GACGTCWACC TCG	33
	(2)	INFOR	MATION FOR SEQ ID NO: 135	
5		(i)	SEQUENCE CHARACTERISTICS:	
•			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
		•	(D) TOPOLOGY: linear	
10			·	
_		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 135	
		RCGHR	CCTTG GGGATAGGTT GTCGCCWTCC ACG	33
15	(2)	INFOR	MATION FOR SEQ ID NO: 136	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
20 ·			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
25		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 136	i
		YCCRG	GCTGR GCCCAGRYCC TRCCCTCGGR YYG	33

	(2)	INFORM	MATION FOR SEQ ID NO: 137	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
5			(B) TYPE: nucleic acid	
-			(C) STRANDEDNESS: single	
		.•	(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
10		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 137	
		BSHRC	CCTCR TTRCCRTAGA GGGGCCADGG RTA	33
15	(2)	INFORM	MATION FOR SEQ ID NO: 138	
12		(4)	SEQUENCE CHARACTERISTICS:	
		(-/	(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
20			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 138	
25		GCCRC	GGGGW GACAGGAGCC ATCCYGCCCA CCC	33
	(0)		ADMINISTRAÇÃO EN MOVELAS	

- 146 -

		(i) SEQUENCE CHARACTERISTICS:	
		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
5		(C) STRANDEDNESS: single	
•		(D) TOPOLOGY: linear	
		(ii) MOLECULE TYPE: DNA	
10	•	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139	
		CCGGGGGTCY GTGGGGCCCC AYCTAGGCCG RGA	33
	(2)	INFORMATION FOR SEQ ID NO: 140	
		(i) SEQUENCE CHARACTERISTICS:	
15		(A) LENGTH: 33 nucleotides	
		(B) TYPE: nucleic acid	
		(C) STRANDEDNESS: single	
		(D) TOPOLOGY: linear	
20		(ii) MOLECULE TYPE: DNA	
		(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140 ATCGATGACC TTACCCAART TRCGCGACCT RCG	33
		NICONSTITUTE	
25	(2)	INFORMATION FOR SEQ ID NO: 141	
		(1) SEQUENCE CHARACTERISTICS:	

- 147 -

			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
5				
		(ii)	MOLECULE TYPE: DNA	
		-	SEQUENCE DESCRIPTION: SEQ ID NO: 141	
		CCCCAT	GAGR TCGGCGAAGC CGCAYGTRAG GGT	33
10				
	(2)	INFORM	ATION FOR SEQ ID NO: 142	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
15			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
20		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 142	
		GCCYCC	WARR GGGGCGCGA CGAGCGGWAT RTA	33
	(2)	INFORM	ATION FOR SEQ ID NO: 143	
25		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	

- 148 -

			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
5		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 143	33
		AACCCG	GACR CCRIGYGCCA RGGCCCIGGC AGC	33
	(2)	INFORM	ATION FOR SEQ ID NO: 144	
10		(i)	SEQUENCE CHARACTERISTICS:	
		,	(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
15				
		(ii)	MOLECULE TYPE: DNA	
			SEQUENCE DESCRIPTION: SEQ ID NO: 144	22
		RTTCCC	TGTT GCATAGTTCA CGCCGTCYTC CAG	33
20				
	(2)	INFORM	ATION FOR SEQ ID NO: 145	
		(i)	SEQUENCE CHARACTERISTICS:	
		(1)	(A) LENGTH: 33 nucleotides	
			(B) TYPE: nucleic acid	
25			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
			(") 102000011 111001	

		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 145	
5		CARRAG	GGAAG AKAGAGAAAG AGCAACCRGG MAR	33
	(2)	INFORM	MATION FOR SEQ ID NO: 146	
		(i)	SEQUENCE CHARACTERISTICS:	
10			(A) LENGTH: 20 nucleotides	
			(B) TYPE: nucleic acid	
		•	(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
15		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 146	
		AGGCAT	AGGA CCCGTGTCTT	20
20	(2)	INFORM	NATION FOR SEQ ID NO: 147	
		(i)	SEQUENCE CHARACTERISTICS:	
			(A) LENGTH: 20 nucleotides	
			(B) TYPE: nucleic acid	
25			(C) STRANDEDNESS: single	
			(D) TOPOLOGY: linear	
		(ii)	MOLECULE TYPE: DNA	
		(xi)	SEQUENCE DESCRIPTION: SEQ ID NO: 147	
		CTTCTT	TGGA GAAAGTGGTG	20

WO 92/19743 PCT/US92/04036

- 150 -

## CLAIMS

1. As a composition of matter, a non-naturally occurring nucleic acid having a non-HCV-1 nucleotide sequence of eight or more nucleotides corresponding to a nucleotide sequence within the hepatitis C virus genome.

- 2. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome is selected from the regions consisting of the NS5 region, envelope 1 region, 5'UT region, and the core region.
- 3. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the NS5 region.
- 20 4. The composition of claim 3 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome is selected from a sequence within sequences numbered 2-22.

15

- 5. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the envelope 1 region.
- 5
  6. The composition of claim 5 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a sequence within sequence numbers 24-32.
  - 7. The composition of claim 1 wherein at least one sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the 5'UT region.
  - 8. The composition of claim 7 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a sequence within sequences numbered 34-51.
    - 9. The composition of claim 1 wherein said nucleotide sequence corresponding to a non-HCV-1 nucleotide sequence within the hepatitis C virus genome corresponds to a sequence in the core region.

10. The composition of claim 9 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence within the hepatitis C virus genome corresponds to a within sequences numbered 53-66.

5

11. The composition of claim 1 wherein said non-naturally occurring nucleic acid has a nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.

10

- 12. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.
- 13. The composition of claim 11 wherein said

  non-naturally occurring nucleic acid has a sequence
  corresponding to a sequence of a second genotype which
  second genotype is defined substantially by sequences
  numbered 7-12 in the NS5 region, 26-28 in the envelope
  1 region, 39-45 in the 5'UT region, and 58-64 in the
  core region.

- 14. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.
- 15. The composition of claim 11 wherein said

  non-naturally occurring nucleic acid has a sequence
  corresponding to a sequence of a fourth genotype which
  fourth genotype is defined substantially by sequences
  numbered 20-22 in the NS5 region, 29-31 in the envelope
  1 region and 48-49 in the 5'UT region.
- 16. The composition of claim 11 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.
- 17. The composition of claim'l wherein said non-naturally occurring nucleic acid is capable of
  25 priming a reaction for the synthesis of nucleic acid to form a nucleic acid having a nucleotide sequence corresponding to hepatitis C virus.

- 18. The composition of claim 1 wherein said non-naturally occurring nucleic acid has label means for detecting a hybridization product.
- 5 19. The composition of claim 1 wherein said non-naturally occurring nucleic acid has support means for separating a hybridization product from solution.
- 20. The composition of claim 1 wherein said

  non-naturally occurring nucleic acid prevents the transcription or translation of viral nucleic acid.
  - 21. A method of forming a hybridization product with a hepatitis C virus nucleic acid comprising the following steps:
- a. placing a non-naturally occurring nucleic acid having a nucleotide sequence of eight or more nucleotides corresponding to a non-HCV-1 sequence in the hepatitis C viral genome into conditions in which hybridization conditions can be imposed said non-naturally occurring nucleic acid capable of forming a hybridization product with said hepatitis C virus nucleic acid under hybridization conditions; and

- b. imposing hybridization conditions to form a hybridization product in the presence of hepatitis C virus nucleic acid.
- 5 22. The method of claim 21 wherein said nucleotide sequence corresponding to a non-HCV-1 sequence in the hepatitis C virus genome corresponds to a sequence within at least one of the regions consisting essentially of NS5 region, envelope 1 region, 5'UT region, and the core region.
  - 23. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within the NS5 region.
- 24. The method of claim 23 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within sequences numbered 2-22.
  - 25. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponds to a sequence within the envelope 1 region.

- 26. The method of claim 25 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence is selected from a sequence within sequences numbered 24-32.
- The method of claim 21 wherein said nucleotide
  sequence corresponds to a non-HCV-1 sequence
  corresponding to a sequence within the 5'UT region.
- 28. The method of claim 27 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence selected from a sequence within sequences numbered 34-51.
- 29. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence corresponding to a sequence within the core region.
- 30. The method of claim 29 wherein said nucleotide sequence corresponds to a non-HCV-1 sequence selected20 from a sequence within sequences numbered 53-66.
  - 31. The method of claim 21 wherein said nucleotide sequence corresponds to a non-HCV-1 nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.

- 32. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.
- 33. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, 39-45 in the 5'UT region, and 58-64 in the core region.
- 15 34. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.
- 35. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.

- 158 -

36. The method of claim 21 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.

5

- 37. The method of claim 21 wherein said hybridization product is capable of priming a reaction for the synthesis of nucleic acid.
- 38. The method of claim 21 wherein said non-naturally occurring nucleic acid has label means for detecting a hybridization product.
- 15 39. The method of claim 21 wherein said non-naturally occurring nucleic acid has support means for separating the hybridization product from solution.
- 40. The method of claim 21 wherein said non-naturally occurring nucleic acid prevents the transcription or translation of viral nucleic acid.
- 41. As a composition of matter, a non-naturally occurring polypeptide corresponding to a non-HCV-1 nucleotide sequence of nine or more nucleotides which sequence of nine or more nucleotides corresponds to a sequence within hepatitis C virus genomic sequences.

- 42. The composition of claim 41 wherein said non-HCV-1 sequence is selected from one of the regions consisting of NS5 region, envelope 1 region, and the core region.
- 5 43. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence corresponds to a sequence in the NS5 region.
- 44. The composition of claim 43 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 2-22.
- 45. The composition of claim 41 wherein said non-HCV-1 sequence corresponds to a sequence in the envelope 1 region.
  - 46. The composition of claim 45 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 24-32.
  - 47. The composition of claim 41 wherein said non-HCV-1 sequence corresponds to a sequence in the core region.
- 48. The composition of claim 47 wherein said non-HCV-1 sequence is selected from a sequence within sequences numbered 52-66.

- 49. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a nucleotide sequence corresponding to one or more genotypes of hepatitis C virus.
- 50. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, and 52-57 in the core region.
- 51. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, and 58-64 in the core region.
- 20 52. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, and 65-66 in the core region.

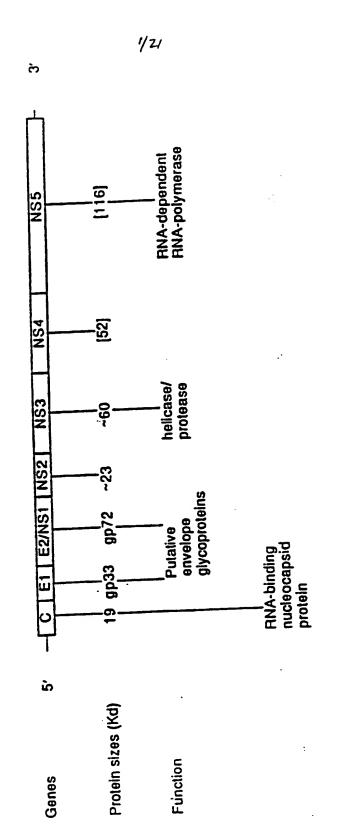
- 53. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.
- 54. The composition of claim 41 wherein said non-HCV-1 nucleotide sequence has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region and 50-51 in the 5'UT region.
- 55. The composition of claim 41 wherein said
  15 polypeptide is capable of generating an immune reaction
  in a host.
  - 56. An antibody capable of selectively binding to the composition of claim 41.
  - 57. A method of detecting one or more genotypes of hepatitis C virus comprising the following steps:
    - a) placing a non-naturally occurring nucleic acid having a nucleotide sequence of eight or more nucleotides corresponding to one or more genotypes of hepatitis C virus under conditions where hybridization conditions can be imposed,

- b) imposing hybridization conditions to form a
   hybridization product in the presence of hepatitis
   C virus nucleic acid; and
- c) monitoring the non-naturally occurring nucleic acid for the formation of a hybridization product, which hybridization product is indicative of the presence of the genotype of hepatitis C virus.
- 58. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a first genotype which first genotype is defined substantially by sequences numbered 1-6 in the NS5 region, 23-25 in the envelope 1 region, 33-38 in the 5'UT region, and 52-57 in the core region.
- 59. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a second genotype which second genotype is defined substantially by sequences numbered 7-12 in the NS5 region, 26-28 in the envelope 1 region, 39-45 in the 5'UT region, and 58-64 in the core region.

PCT/US92/04036

- 60. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a third genotype which third genotype is defined substantially by sequences numbered 13-17 in the NS5 region, 32 in the envelope 1 region, 46-47 in the 5'UT region and 65-66 in the core region.
- 61. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fourth genotype which fourth genotype is defined substantially by sequences numbered 20-22 in the NS5 region, 29-31 in the envelope 1 region and 48-49 in the 5'UT region.
- 15 62. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence of a fifth genotype which fifth genotype is defined substantially by sequences numbered 18-19 in the NS5 region.
  - 63. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 67-145.

- 64. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 69, 71, 73 and 81-99 to identify Group I genotypes in the core and region of the HCV genome.
- 65. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 70, 72, 70 and 100-118 to identify Group II genotypes in the core and envelope regions of the HCV genome.
- 66. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence corresponding to a sequence numbered 77 to identify Group III genotypes in the 5' UT region of the HCV genome.
- 67. The method of claim 57 wherein said non-naturally occurring nucleic acid has a sequence numbered 79 to identify Group IV genotypes in the 5' UT region of the HCV genome.



SUBSTITUTE SHEET

2/2-1

Fig. 2a

. ; :

NS5 REGION

SEQUENCE IN MIMBER		U) 31 2)	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 11 10 10 11 11 11	11 15 11 11 11 11 11 11	## ## ## ## ## ## ## ## ## ## ## ## ##	11 11 12 13 14 14 14 14 14 15 14 15 15 16 16 17 18 18 18 18 18 18 18 18 18 18 18 18 18	# # # # # # # # # # # # # # # # # # #	
	13		1 CTCCACAGTC A 1 CTCCACAGTC A 1 CTCCACAGTC A 1 CTCCACAGTC A 1 CTCTACAGTC A 1 CTCTACAGTC A 1 CTCTACAGTC A	AGTC ACTGAGAGCG ACATCCGTAC GGAGGAGGCA ATCTACCAAT AGTC ACTGAGAGCG ACATCCGTAC GGAGGAGGCA ATCTACCAAT AGTC ACTGAGAGCG ACATCCGTAC GGAGGAGGCA ATCTACCAAT AGTC ACTGAGAACG ACATCCGTAC GGAGGAGGCA ATCTACCAAT AGTC ACTGAGAACG ACATCCGTAC GGAGGAGGCA ATCTACCAAT AGTC ACTGAGAGCG ATATCCGTAC GGAGGAGGCA ATCTACCAAT	ACATCCGTAC ACATCCGTAC ACATCCGTAC ACATCCGTAC ATATCCGTAC ATATCCGTAC	GGAGGAGGCA ATCTACCAAT GGAGGAGGCA ATCTACCAAT GGAGGAGGCA ATCTACCAAT GGAGGAGGCA ATCTACCAAT GGAGGAGGCA ATCTACCAAT GGAGGAGGCA ATCTACCAAT	ii — — —		CGACCCCAA GGACCCCCAA GGACCCCCAA GGACCCCCAA GGACCCCCAA
11 10 10 12	e e n e n e n e n e n e n e n e n e n e	"	CTCCACGGTC CTCAACGGTC CTCAACGGTC CTCAACGGTC CTCCACGGTC	ACTGAGATG ACTGAGATG ACCGAGATG ACTGAGATG ACTGAGATG		TGAGGAGTCA TGAGGAGTCA TGAGGAGTCA CGAGGAGTCG TGAGGAGTCG	1 1 M 1 M 1 M 1 M 1 M 1 M 1 M 1 M 1 M 1	GTTGTGACTT GTTGTGACTT GTTGTGACTT GTTGTGACTT GTTGTGACTT	gaa gag gag gaa gaa gaa
13 13 14 15 16 17	8			ACTGAGAGAG ACATCAGAAA ACGTAAAAGG ACATCACAT ACAGAGAGG ACATCAGAAA ACGGAGAGGG ACATCAGAAA	ACATCAGAAC ACATCACATC ACATCAGAAC ACATCAGAAC ACATCAGAAC	ដើមិបិបិនី	I KKKKK!	ATATACCGAG CCTGCTCCCT ATCTACCAGT CCTGTTCACT ATCTATCTGT CCTGCTCACT ATCTATCTGT CCTGTTCACT	GCCTGAGGAG GCCCGAGGAG GCCTGAGGAG GCCTCAGGAG
19		11	น	TCGACCGIT ACCGAACATG ACATAATGAC	ACATAATGAC	TGAAGAGTCT TGAAGAGTCC	144 !	CATTGTACTT CATTGTACTT	
======================================	GIV	#	CTCTACTGTC CTCGACTGTC CTCGACTGTC	CTCTACTGTC ACTGAACAGG ACATCAGGGT G CTCGACTGTC ACTGAACAGG ACATCAGGGT G CTCAACTGTC ACTGAACAGG ACATCAGGGT G	ACATCAGGGT ACATCAGGGT ACATCAGGGT	GGAAGAGGAG GGAAGAGGAG GGAAGAGGAG	GGAAGAGGAG ATATACCAGT GCTGTAACCT GGAAGAGGAG ATATACCAAT GCTGTAACCT GGAAGAGGAG ATATACCAAT GCTGTAACCT	GCTGTAACCT GCTGTAACCT GCTGTAACCT	CTCTACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG ATATACCAGT GCTGTAACCT TGAACCGGAG CTCTACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG ATATACCAAT GCTGTAACCT TGAACCGGAG CTCAACTGTC ACTGAACAGG ACATCAGGGT GGAAGAGGAG ATATACCAAT GCTGTAACCT TGAACCGGAG
91 91 91 91 91 91	11 11 11 11 11 11 11	11 12 14 13 13 13 13	n	0 6 6 7 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	n H H H H H H H H H H H H H H H H H H H	r t t t t 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	 	i i	

3/21

NSS REGION - (2/5)

SEQUENCE	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1							
œ	GENOTYPE			1		11 11 11 11 11 11 11	. 经银行日十十十十十年日日数11811180分列内线设计的共同时间间对加速设置的设计下下,1986年1987年1988年1988年1987年1988年1987年1988年1988		11 11 11 11 11 11 11
## ## ## ## ## ## ## ## ## ## ## ## ##				AAGTC CC	TCACCGAG 7	CCTCACCGAG AGGCTTTATG	TIGGGGGCCC ICTTACCAAT		TCAAGGGGGG
-4 r	5 5	1.5	GCCCCATGG CCATC		TCACTGAG 1				TCAAGGGGGG
۷ ۳	; ; ;	11	GCCGCGTGG CCAT		CCTCACTGAG 1	_			TCAAGGGGGG
· <	19	7.1	CCCCCCTCC CCAT	CCATCAAGTC CC	TCACTGAG				TCAAGGGGGG
יטי	61	11		CCATCAAGTC CC	TCACCGAG		TCGGGGGCCC TCT		TCAAGGGGGG
•	GI	7.1	GCCCGTGTGG CCAT	CCATCAAGTC CC	TCACTGAG	CCICACIGAG AGGCIFIAIG		TCIFACCAAL	ו ו
	11 05 01 01 01 11 01 01	1) 31 11 11 11	计设计设计设计设计 计自己转移 医多种性 医多种性 医多种性 医多种性 医多种性 医多种性 医多种性 医多种性	0 D ## ## ## ## ## ## ## ## ## ## ## ## #	H H H H H H H H H H H H H H H H H H H			13	1   1   1
7	GII	71	9	AAGGTC GC	TCACAGAG	GCTCACAGAG CGGCTCTATG TCGGGGGTCC	TCGGGGGTCC TAT		TCCAAAGGGC
. cc	1	71	GCCAGACAAG CCATAAGGTC						TCAAAGGGC
		71	GCTAGACAGG CCATAAGGTC		GCTCACAGAG	CGGCTTTATA			TCAAAGGGGC
, -		11	GCCAGGCAGG, CCATAAGGTC		GCTCACCGAG	CGACTITATA			TCAAAGGGC
3 5		7.1	GCCAGACAGG CTATAAGGTC		GCTCACAGAG	CGGCTGTACA			TCAAAGGGC
11		7.1	GCCAGACAGG CTATAAGGTC		GCTCACAGAG	CGGCTTTACA	rceeggerce cer	CCTGACTAAT T	
•		1	112 113 113 113 113 113 113 113 113 113	- 11				0 0 10 10 10 10 10	11 11 11 11 11 11 11 11 11 11 11 11 11
H H H H H T H T H			GCTCACATTG		CTGACTGAG	AGGCTCTACG	GCTGACTGAG AGGCTCTACG TGGGAGGGCC CAT	GTTCAAC	CATGITCAAC AGCAAGGGCC
7 -		7.1			ACTGACTGAG		TAGGGGGGCC	GACAAAC A	CATGACAAC AGCAAGGCC
T 1		7.1			ACTGACTGAG	AGACTGTACG	TAGGGGGGCC	CATGACAAAC	AGCAAGGGGC
91		71			ACTGACTGAG	AGGCTGTACG	TAGGGGGGCC	CATGACAAC	AGCAAAGGGC
17		11	GCTAGAACTG CTAT	CTATCCACTC G	CTCACTGAG	GCTCACTGAG AGACTCTACG	TAGGAGGGCC CAT	CATGACAAAC	ACCAACCCAC
() () () () () () ()	## 11	11 11 11	11 11 11 11 11 11 11 11 11 11	6 6 11 11 11 11 11	11 11 11 11 11 11 11		THE THE CANADA A STATE OF THE CONTROL OF THE CANADA AND A STATE OF THE		1
18		71	GCGCGTGTGG CAAT	PACGGIC A	CICACCCAA	CCCTGTACE	SCGCGTGTGG CAATACGGTC ACTCACCCAA CGCCTGTACT GTGGAGGCCC CALGLALAAC AGCAAGGGC	ICIALAAC A	
19		7.1	GCACGCGCGG CAA:	CAATACGGTC A	CTCACCCAA	CGCCTGTACT	ACTCACCCAA CGCCTGTACT GTGGAGGCCC CAIGIAIAAC AGCAAGGCC	TOTATANC	
11 21 21 21 21 21 21	11 11 11 14 15 19 19	11 13 14 11 11	11 11	11 11 11 11 11	8				)   
20	S	7.1	GCCAGGAAAG TGATCTCCTC	rerecte e	CTCACGGAG	CGGCTTTACT		TOTTCHAL	ACCARGGGG
2.1		71	GCCAGGAAAG TGATCTCCTC		CTCACGGAG	CCTCACGGAG CGGCTTTACT	GCGGGGCCC TAIGITCAAL AGCAAGGGG	TGIICAAL	AGCAAGGGGG
22		11	GCCAGGAAAG TGATCTCCTC	rerecte e	CTCACGGAA	CGGCTTTACT	CCTCACGGAA CGGCTTTACT GCGGGGCCC TATGLTCARC	דפדורשער	100000000000000000000000000000000000000
11 11 10 11 11 11 11 11	15 10 11 11 11 11 11 11 11 11 11 11 11	11 11 11 11	11 11 11 11 11 11 11 11 11 11 11 11 11		11 15 15 15 16 11 11 11	11 11 12 13 14 14 14	i i i i i i i i i i i i i i i i i i i	1 1 1 1 1 1 1	 

11	
13 13	
11 11 11	
11 11 11	
11    	
H	
ii ii	
ti U U	
.ii !!	
N 11	
11	
1) 1) 11	
11	
i) i1	
11	
13  1  1	
  11	
11	
II II II	
t t	
1	
1	

NSS REGION - (3/5)

61 41 11 01 11 61 61 61 61 71 81	13 14 18 19 19	n H H U	99 88 88 88 88 88 88 88 88 88 88 88 88 8	11 11 11 11 11 11 11	11 13 11 11 11 11 11 11	11 41 41 41 41 41 41 41 41 41	11 11 11 11 11 11 11 11	## ## ## ## ## ## ## ## ## ## ## ## ##	11 10 11 11 10 10 10 11 11 11
SEQUENCE									
ID NUMBER	GENOTYPE		0	11 11 11 11 11 11 11 11 11 11 11 11 11	## ## ## ## ## ## ## ## ## ## ## ##	11 13 13 14 14 14 15 15 16 16 17 18	0 7 0 15 7 15 16 18 18	H H H H H H H H H H H H H H H H H H H	
0 4 11 11 11 11 11 11 11 11	11 11 11	11	11 C 11 <	TOUR DE TOUR	S CCCCCCA G	GCGCCCTACT G	GACAACTAGC T	TGTGGTAACA C	CCCTCACTTG
-	3	167	STATE SOCIOLOGICAL STATE					TGTGGTAACA C	CCCTCACITG
~		141							CCTCACTTG
e		141	AGAACTGCGG CTAC						
		141	AAAACTGCGG CTA7	-	Teccecce				מוניאניאניים
. ر.		141	AAAACTGCGG CTAT	CTATCGCAGG TG			GACAACTAGC		
9		141	_	CTACCGCAGG TG	TGCCGCGCAA	- 11	GACGACTAGC 1	TGTTOTTOTTOTTOTTOTTOTTOTTOTTOTTOTTOTTOTT	
0) 10 11 15 11 11 11 11	U 01 11 11			ranterra Tarcacaca academ	CCCCCCCCA	CGGCGTGCT (			CCCTCACATG
_	119			DI REJUBULKED	TGCGGGGCA	SCGGTGTGCT (	GCGGTGTGCT GACGACTAGC TGCGGTAATA		CCCTCACATG
80		T . T	-			GCGGCGTACT	GACGACCAGC TGCGGTAATA		CCCTTACATG
6		141	-	-		GCGCCTGCT GACGACTAGC	GACGACTAGC		CCCTCACATG
10		141			くれていている	こうじょうこうこう ていこうこうごうしょ		TGCGGTAACA	CCCTCACATG
11		141			פרפרפראי		. טטעבטעטטעט	TCCCCTAATA	CCTCACATG
12		141	AGAACTGCGG CTA	CTATCGCCGG TG	SCCCCCAA	TGCCGCGCAA GCGGCGIGCI GACGACIAGC	TOOTOTO OF THE PERSON TOO TOO TOO TOO TOO TOO TOO TOO TOO T		
11 15 15 10 10 11 11 17	11 10 11 11 11 11 11	11 11 11 11 11						ATGGGGAACA	ບ
13	CIFI	141			GCCCCCA				CACTCACGTG
14		141		GIACAGGCGT TO				WINDOOT WIND	CGCTCACGTG
15		141		GIACAGGCGT TGCCGCGCGA				ATCCCTAACA	CACTCACGTG
16		141		GTACAGGCGT TGCCGCGCGA	GCCGCGCGA	GCGGAGIGCI	つりないしょしてい	ATGGGGAATA	CCATGACATG
11		141	AAICCIGCGG IIA	TTACAGGCGT I	TGCCGCGCCA	GUGGGGTCII	これによってある。		
11 11 11	11 13 13 11 11	B + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 +	1) 1)	E VUVEULE	GCCGCGCCA	GCGGCGTCTT	CACCACTAGT	ATGGGCAACA	CCATGACGTG
18	CA CA	141	AACANIGIGG TINICGITION TOCCOMMENT OF TACTACT ATGGCAACA	* *************************************	***************************************		CACCACCAGT	ATGGGCAACA	CCATGACGTG
19		141	AACAATGTGG TE	ACCGTAGA I	פררפרפררא			H H H H H H H H H H	11 17 11 11 11 11 11 11 11 11
1) 1) 1) 1) 1) 1) 1)	11 11 11 11 11 11 11 11 11 11 11 11 11	01 11 12 13 13 13 13	sessessessessessessessessessessessesses		マルンジルジンンン	GTGGAGTCCT	GCCTACCAGC	TTCGGCAACA	CAATCACTIG
20	GIV	141	CCCAGIGIGG 11	AICCCCAL A	くまつりまりつり	TOTAL DESIGNATION OF THE PARTY	IIAICGCCGI IGCCGIGGAGTTTT GFTACTAGG TTCGGCAACA	TTCGGCAACA	CAATCACTTG
21		141	CCCAGTGTGG TIATCGCCCGT TGCCGTGCTA GIGGAGTICA GCCTACCCCC TTCGGCCAACA	ATCGCCGE 1	ופררפופרזא	TOTAL TOTAL TOTAL	CCTACCAGG	TTCGGCAACA	CAATCACTIG
22		141	CCCAGIGIGG II	Arceccer 1	ופנרפזפרנא	OTOCKOTO.			11 12 13 14 14 15 16 17 17 18 18 18 18 18 18 18 18 18 18 18 18 18
11 11 10 10 11 10 11	11 11 11 11 11 11 11 11 11 11	11 11 11 11 11		11 11 11 11 11 11 11	H H H H H H H H H H H		1 1 1 1 1 1 1 1 1		

TICAAG GCCGGGCAG CCTGTCGAGC CGCAGGCTC CAGGACTGCA CCATGCTCGT ( TCAAG GCCCGGGCAG CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTCGT ( TTAAG GCCCGGGCAG CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTCGT ( TTAAG GCCCGGGCAG CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTCGT ( TTAAG GCCCGGGCAG CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTCGT ( TTCAAG GCCCAAGCGC CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTCGT ( TTCAAG GCCCTGCAGC CCTGTCGAGC CGCAGGGCTC CAGGACTGCA CCATGCTCGT ( TTCAAG GCCCTGCAGC CTGTCGAGC CGCAGGCTC CAGGACTGCA CGATGCTCGT ( TTGAAG GCCTCTGCAGC CTGTCGAGC TGCCAAGCTC CAGGACTGCA CGATGCTCGT ( TTGAAG GCCTCTGCAGC CTGTCGAGC TGCCAAGCTC CAGGACTGCA CGATGCTCGT ( TTGAAG GCCTCTGCAGC CTGTCGAGC TGCCAAGCTC CAGGACTGCA CGATGCTCGT ( TTGAAG GCCTCTGCAGC CTGCTAGAGC TGCCAAGCTC CAGGACTGCA CGATGCTCGT ( TTGAAG GCCTTGCAGC CTGCTAGAGC TGCCAAGCTC CAGGACTGCA CGATGCTCGT ( TTGAAG GCCTTGCAGC CTGCGAGCTC CAGGACTGCA CGATGCTCGT ( TTGAAA GCCTTGCAGC CTGCGAGCTC CAGGACTGCA CGATGCTGT ( TTGAAA GCCTTGCAGC CTGCCAAGCT GTGCACCCT CAATGCTGGT ( TTGAAA GCCTTGCAGC CTGCCAAGCT GTGCACCCT CAATGCTGGT ( TTGAAA GCCTTGCAGC CTGCCAAAGCT GTGCACCCT CAATGCTGGT ( TTCAAA GCCTTGCAGC CTGCAAAGCC CGCGGGATT GTTGCTCCCA CCATGTTGGT ( TTCAAA GCCTTGCAGC CTGCAAAGCC CGCGGGCATT GTTGCTCCCA CCATGTTGGT ( TTCAAA GCCTTGCAG CTGCAAAGC CGCGGGCATT GTTGCTCCCA CCATGTTGGT ( TTCAAA GCCTTGCAG CTGCAAAGC CGCGGGCATT GTTGCTCCT CAATGCTGGT ( TTCAAA GCCTTGCAG CTGCAAAGC CGCGGGCATT GTTGCTCCT CAATGCTGGT ( TTCAAA GCCTTGCAG CTGCAAAGC CGCGGGCATT GTTGCTCCCA CCATGTTGGT ( TTCAAA GCCTTGCAG CTGCAAAGC CGCGGGCATT GTTGCTCCA CCATGTTGGT ( TTCAAA GCCTTGCAG CTGCAAAGC CGCGGGCATT GTTGCTCCT CAATGCTGGT ( TTCAAAA GCCTTGCAG CTGCAAAGC CGCGGGCATT GTTGCTCCCA CCATGTTGGT ( TTCAAAA GCCTTGCAG CTGCAAAGC CGCGGGCATT GTTGCTCCA CCATGTTGGT ( TTCAAAA GCCTTGCAG CTGCAAAGC CGCGGGCATT GTTGCTCCA CCATGTTGT ( TTCAAAA GCCTTGCAG CTGCAAAGC CGCGGGCATT GTTGCTGCA CCTTGTTGT ( TTCAAAAAAA GCCTTGCAGCA CCTGCAAAGCT CCAGGCTCCA CAATGCTGCA C
SERVITYEE

Fig. 2e

NSS REGION - (5/5)

reenresnennes de constant de mange de m SEQUENCE

GENOTYPE

ID NUMBER

1

ID NUMBER GENOTYPE	ı	11 11 11 11 11	. 6. 11 6. 6. 6. 6. 6. 6. 6. 6. 6. 6. 6. 6. 6.	11 13 14 14 15 15 16 17 18 18 18 18 18 18 18 18 18 18 18 18 18		11 12 13 13 14 14 14 14 14 14 14 14 14 14 14 14 14	0 0 11 0 11 0 0	11 11 11 11 11 11 11 11	0) 9) 11 0) 6)
11	!	281 281 281 281 281 281	GACTTAGTCG TTATCTGTGA AAGCGCGGGG GTCCAGGAGG ACGCGGCGAG CCTGAGAGCC GACTTAGTCG TTATCTGTGA AAGTGCGGG GTCCAGGAGG ACGCGGCGAG CCTGAGAGCC GACTTGGTCG TTATCTGTGA GAGTGCGGGG GTCCAGGAGG ACGCGGCGAG CCTGAGAGCC GACTTAGTCG TTATCTGTGA GAGTGCGGGA GTCCAGGAGG ACGCGGCGAA CTTGAGAGCC GACTTAGTCG TTATCTGTGA AAGTGCGGGA GTCCAGGAGG ATGCAGCGGAA CCTGAGAGCC GACTTAGTCG TTATCTGCGA AAGTGCGGGG GTCCAGGAGG ACGCGCGAA CCTGAGAGCC GACTTAGTCG TTATCTGCGA AAGTGCGGGG GTCCAGGAGG ACGCGCCGAA CCTGAGAGCC	TTATCTGTGA AAGCGCGGGG TTATCTGTGA AAGTGCGGGG TTATCTGTGA GAGTGCGGGGA TTATCTGTGA AAGTGCGGGGA TTATCTGTGA AAGTGCGGGGA TTATCTGCGA AAGTGCGGGGA	TTATCTGTGA AAGCGCGGGG GTCCAGGAGG TTATCTGTGA AAGTGCGGGG GTCCAGGAGG TTATCTGTGA GAGTGCGGGG GTCCAGGAGG TTATCTGTGA GAGTGCGGGA' GTCCAGGAGG TTATCTGTGA AAGTGCGGGG GTCCAGGAGG	GTCCAGGAGG GTCCAGGAGG GTCCAGGAGG GTCCAGGAGG GTCCAGGAGG	ACGCGCCGAG ACGCGCCGAG ACGCGCCGAG ACGCGCCGAA ACGCGCCGAA ACGCGCCGAA	CCTGAGAGCC CCTGAGAGCC CCTGAGAGCC CTTGAGAGCC CCTGAGAGCC CCTGAGAGCC	H II II
10 11 12	11 11 11 11 11 11 11 11 11 11 11 11 11	281 281 281 281 281 281 281	JI .	TTATCTGTGA TTATCTGTGA TTATCTGTGA TTATCTGCGA TTATCTGTGA	TTATCTGTGA AAGCGCGGGG TTATCTGTGA AAGCGCGGGA TTATCTGCGA GAGCGCGGGA TTATCTGTGA GAGCGCGGGA	SACCTIGICG TIAICIGIGA AAGCGCGGGG AACCAAGAGG ACGCGGCAAG CCIACGAGCCGACCTGTCTCG TIAICIGIGA AAGCGCGGGA ACCCAGGAGG AIGCGGCGAG CCTACGAGTCGACTIGICG TIAICIGIGA AAGCGCGGGA ACCCAGGAGG ACGCGGCGAA CCTACGAGTGGACCTIGICG TIAICIGIGA GAGCGCGGGA ACCCAAGAGG ACGCGGCGAG CCTACGAGTGGACCTIGICG TIAICIGIGA GAGCGCGGGA ACCCAAGAGG ACGCGGGGAG CCTACGAGTGGACTIGICG TIAICIGIGA GAGCGCGGGA ACCCAAGAGG ACGCGGCGAG CCTACGAGTGGACTIGICG TIAICIGIGA GAGCGCGGGA ACCCAAGAGG ACGCGGCGAG CCTACGAGTGGACTIGICG TIAICIGIGA GAGCGCGGG ACCCAAGAGG ACGCGGCGAG CCTACGAGTGGACTIGICG TIAICIGIGA GAGCGCGGG ACCCAAGAGG ACGCGGCGAG CCTACGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAG	AACCAAGAGG ACGCGGCAAG CCTACGAGCC ACCCAGGAGG ATGCGGCGAA CCTACGAGTC ACCCAAGAGG ACGCGGCGAA CCTACGAGTC ACCCAAGAGG ACGCGGCGAG CCTACGAGTC ACCCAAGAGG ACGCGGCGAG CCTACGAGTC	11	11 11 10 11 11
13 13 14 15 15 17	H I I I I I I I I I I I I I I I I I I I	281 281 281 281 281 281	11	TCATCTCAGA TCATCTCAGA TCATCTCAGA TCATCTCAGA TCATCTCGGA	TCATCTCAGA AAGCCAGGGG TCATCTCAGA GAGTCAGGGG TCATCTCAGA GAGTCAGGGG TCATCTCAGA GAGTCAAGGG	TCATCTCAGA AAGCCAGGG ACTGAGGAGG ACGAGCGGAA CCTGAGAGCT TCATCTCAGA GAGTCAAGGG GCTGAGGAGG ACGAGCAGAA CCTGAGAGTC TCATCTCAGA GAGTCAGGG GTCGAGGAAG. ATGAGCGGAA CCTGAGAGTC TCATCTCAGA GÀGTCAAGGG GTCGAGGAGG ATGAGCGAAA CCTGAGAGCT TCATCTCGGA GAGCGAAGGT AACGAGGAGG ACGAGCGAAA CCTGAGAGCT	ACGAGCGGAA ACGAGCGGAA ATGAGCGGAA ATGAGCGGAA ACGAGCGAAA	11	, n , n n u
======================================	# 25 # 25 # # # # # # # # # # # # # # # # # # #	li .	GATCTTGTGG	CCATTIGCGA	GAGCCAAGGG	ACGCACGAGG ACGCACGAGG	ATAAAGCGAG ATGAAGCGTG	GATCTTGTGG CCATTTGCGA GAGCCAGGGG ACGCACGAGG ATAAAGCGAG CCTGAGAGCC ACCTTGGTGG CCATTTGCGA GAGCCAAGGG ACGCACGAGG ATGAAGCGTG CCTGAGAGTC	21 11 11 11 11
20 21 22	GIV	281 281 281	281 GATCTGGTCG TGGTGGCTGA GAGTGATGGC GTCGACGAGG ATAGAGCAGC CCTGAGAGCCC 281 GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCGACGAGG ATAGAACAGC CCTGCGAGCC 281 GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCAATGAGG ATAGAGCAGC CCTGGGAGCC 281 GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCAATGAGG ATAGAGCAGC CCTGGGAGCC	TGTGGCTGA TGTGGCTGA TGTGGCTGA	GAGTGATGGC GAGTGATGGC GAGTGATGGC	GATCTGGTCG TGGTGGCTGA GAGTGATGGC GTCGACGAGG ATAGAGCAGC GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCGACGAGG ATAGAACAGC GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCAATGAGG ATAGAGCAGC	ATAGAGCAGC ATAGAACAGC ATAGAGCAGC	GATCTGGTCG TGGTGGCTGA GAGTGATGGC GTCGACGAGG ATAGAGCAGC CCTGAGAGCC GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCGACGAGG ATAGAACAGC CCTGGGAGCC GATCTGGTTG TGGTGGCTGA GAGTGATGGC GTCAATGAGG ATAGAGCAGC CCTGGGAGCC	11 15 10 11
11 11 11 11 11 11 11 11	1 1 1 1 1 1 1 1 1 1	;   							

340 TOTAL

7/2/

BENVETPE  GENOTIFE  GI 1 GACGGGTTG GTANTGGTC GATCCCANA GCCATCTTGG ACTTANGCC  1 GACGGGTTG GTANTGGTC GATCCCCANA GCCATCTTGG ACTTANGCC  1 ACCGGGTTG GTANTGGTC GATCCTCCG GATCCCANA GCCATCTTGG ACTTANTCGC  GI 1 ACCGGGTTG GTANTGGTC GATCCTCCG GATCCCCANA GCCATCTTGG ACTTANTCGC  GI 1 ACCGGGTTG GTANTGGTC GATCCTCCG GATCCCCANA GCCATCTTGG ACTTANTCGC  GTANTGGTTGC GTANTGGTTCG GATCCTCCANA GCCATCTTGG ACTTANTCGC  GTANTGGTTANGC GATCCTCCC GATCCTCCCANA GCCATCTTGG ACTTANTCGC  GTANTGGTTANG GTGGTTGGCC ACTTANTCCG GATCCCCCANA ACCTTGTTCG ACTTANTANGC  1 TOTGGGTANG GTGGTTGGCC ACTTCCTCCC GATCCCCCANA ACCTTGTTCG ACATCANAGC  GTANTGGTANG GTGGTTGGCC ACTTCCTCCC GATCCCCCANA ACCTTGTTCG ACATCANAGC  GTANTGGTANG GTGGTTGGCC ACTTCCTCCC GATCCCCCANA ACCTTGTTCG ACATCANAGC  GTGTGCTCAC TGGGGAATCC TGGCGGGCANA ACCTTGTTCG ACATCANAGC  GTGTGCTCAC TGGGGAATCC TGGCGGGCANA ACCTTATTC  GTGTGCCCCAC TGGGGAATCC TGGCGGGCANA ACCTTATTC  GTGTGCCCACA TGGGGAATCC TGGCGGGCANA ACCTTATTC  GTGTGCCCACA TGGGGAATCC TGGCGGGCANA ACCTTATTATC  GTGTGCCCACA TGGGGAATCC TGGCGGGCANA ACCTTATTATC  GTGTGCCCACA TGGGGAATCC TGGCGGGCANATCT  GTGGGCCCACA TGGGGGAATCC TGGCGGGCANATCT  GTGGGCCCCACA TGGGGGAATCC TGGCGGGCANATCT  GTGGGCCCACA TGGGGGAATCC TGGCGGGCANATCT  GTGGGCCCACA TGGGGGAATCC TGGCGGGCANATCT  GTGGGCCCACA TGGGGGAATCC TGGCGGGCANATCT  GTGGGCCCACA TGGGGCCCAC TGGCGGGCANATCT  GTGGGCCCACA TGGGGGCACA TGGCGGGCANATCT  GTGGGCCCCACA TGGGGGCACA TGGCGGGCANATCT  GTGGGCCCACA TGGGGGCCCAC TGGCGGGCANATCT  GTGGGCCCACA TGGGGCCCAC TGGCGGGCANATCT  GTGGGCCCACA TGGGGCCCAC TGGCGGGCANATCT  GTGGCCCCACA TGGGGCCCAC TGGCGGCCAC TGCCGGGCCACACACCACAC				Fig. 3
GENOTYPE  GII 1 G  1 A  1 B  GII 1 G  GIV 1 T  T  GIV 1 T  GIV 61				
GI I G 1 A 1 B 1 B 1 B 1 B 1 B 1 B 1 T 1 T 1 T 1 T 1 T 1 T 1 T 1 T	SEQUENCE ID NÜMBER GE	ERNOTYPE	11 11 11 11 11	
6 GII 1 G 8 1 G 8 1 G 9 GIV 1 T 10 1 T 11 T 12 GIII 1 T 23 GI 61 T 24 GI 61 T 25 GI 61 T 26 GI 61 T 27 GI 61 T 28 GI 61 T 29 GIV 61 G 30 GI 61 T 31 GI 61 T 32 GI 61 T 33 GI 61 T 61	0 H H	:: :: :: :: :: :: :: :: :: :: :: :: ::		GCGTTG GTAATGGCTC AGCTGCTCCG GATCCCACAA GCCATCTTGG ACATGATCGC GCGTTG GTGGTAGCTC AGCTACTCCG GATCCCACAA GCCATCATGG ACATGATCGC GCGCTG GTAGTAGCTC AGCTGCTCAG GGTCCCGCAA GCCATCGTGG ACATGATCGC
29 GIV 1 T 30 1 1 T 31 GIII 1 T 23 GI 61 7 24 61 7 25 GII 61 61 7 26 GII 61 61 7 27 61 61 7 28 61 7 29 GIV 61 61 61 81 81 81 81 81 81 81 81 81 81 81 81 81	6 7 8	0 1 I E E E E E E E E E E E E E E E E E E	#	GCCCTA GIGGTATCGC AGTTACTCCG GATCCCACAA GCCGTCATGG ATATGGTGGC GCCCTA GIGGTGTCGC AGTTACTCCG GATCCCACAA AGCATCGTGG ACATGGTGGC GCCCTA GIGGTGTCGC AGTTACTCCG GATCCCGCAA GCTGTCGTGG ACATGGTGGC
32 GIII 11 1 23 GI 61 2 24 61 7 25 GII 61 7 27 61 61 2 29 GIV 61 30 61 31 61	29 30 31	11 11		GGTATG GIGGIGGCGC ACGICCIGCG TITGCCCCAG ACCITGIICG ACAIAAIAGC GGTAIG GIGGIAGCAC ACGICCIGCG ICIGCCCCAG ACCITGIICG ACAIAAIAGC GGTAIG GIGGIGGCGC AAGICCIGCG ITIGCCCCAG ACCITGIICG ACGIGCIAGC
23 GI 61 7 24 61 7 25 GII 61 2 27 GIV 61 29 GIV 61 30 61 31 61	32	GIII	- :	
26 GII 61 61 28 61 61 61 61 61 61 61 61 61 61 61 61 61	23 24 25 25		= <del>-</del>	TGCTCAC TGGGGAGTCC TGGCGGGCAT AGCGTATTTC AGCCCAC TGGGGAGTCC TGGCGGGCAT AGCGTATTTC TGCCCAC TGGGGAGTCC TAGCGGGCAT AGCGTATTTT
29 GIV 61 30 61 31 61	26 27 28	# E # E # E # E # E # E # E # E # E # E	61 61 61	GGCCCAC TGGGGAGTCC TGGCGGGCCT TGCCTACTAT GGCCCAC TGGGGAGTCC TGGCGGGCCT TGCTTACTAT GGCCCAC TGGGGAATCC TAGCGGGTCT TGCCTACTAT
32 GIII 61	29 30 31	8 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	61 61 61	GGCCCAT TGGGGCATCT TGGCGGGCTT GGCCTATTAC GGCCCAT TGGGGCATCT TGGCAGGCCT AGCCTATTAC GGCCCAT TGGGGCATCT TGGCGGGCCT GGCCTATTAC
		11 11	61	11 11

100 Total

Fig. 4a

5'UT Region

ID NUMBER GENOTYPE  33 GI 1 G 34 1 G 35 1 G G 36 1 G 37 1 G 38 1 G G 40 1 G 41 1 G 42 1 G 43 1 G 44 1 G 44 1 G 45 1 G 45 1 G 47 1 G 48 1 G 48 GIII 1 G 47 1 G 47 1 G 48 GIII 1 G 48 GIII 1 G	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC CGCGCCCCCCCC	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT  GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT  GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT  GTTAGTATGA GTGTCGTGCA GCCTCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT  GTTAGTATGA GTGTCGTGCA GCCTCCCAGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT  GTTAGTATGA GTGTCGTGCA GCCTCCCAGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT  GTTAGTAGTATGA GTGTCGTGCA GCCTCCCCCCCCCCC
	GTTAGTATGA GTGTCGTGCA GCCTCC	AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT
34 35 36 37 39 40 41 42 43 44 46	GTTAGTATGA GTGTCGTGCA GCCTCC	AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT AGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT AGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT AGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT ATAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
35 36 37 38 40 41 42 44 45 44 45	GTTAGTATGA GTGTCGTGCA GCCTCC	AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT
36 37 38 39 40 41 42 43 44 45 46 47	GTTAGTATGA GTGTCGTGCA GCCTCC	AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT
37 38 39 40 41 42 43 44 45 45	GTTAGTATGA GTGTCGTGCA GCCTCC	AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT
38 39 40 41 42 43 44 45 46	GTTAGTATGA GTGTCGTGCA GCCTCC	AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT  ### CONTROL CGGGAGAGCC ATAGTGGTCT  ### AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT  ### AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT  #### AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT  #### AGGA CCCCCCTTCC CGGGAGAGCC ATAGTGGTCT  ###############################
39 40 41 42 43 44 45 46	GTTAGTATGA GTGTCGTGCA GCCTCC GTTAGTATGA GTGTCGTGCA GCCTCC GTTAGTATGA GTGTCGTGCA GCCTCC GTTAGTATGA GTGTCGTGCA GCCTCC	AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT
39 40 41 42 43 44 45 46	GTTAGTATGA GTGTCGTGCA GCCTCC GTTAGTATGA GTGTCGTGCA GCCTCC GTTAGTATGA GTGTCGTGCA GCCTCC GTTAGTATGA GTGTCGTGCA GCCTCC	AGGA CCCCCCTCC CGGGAGGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT
40 41 42 43 44 45 ==============================	GITAGIAIGA GIGICGIGCA GCCTCC GITAGIAIGA GIGICGIGCA GCCTCC GITAGIAIGA GIGICGIGCA GCCTCC	AGGA CCCCCCTCC CGGGAGAGCC ATAGTGGTCT AGGA CCCCCCCTCC CGGGAGAGCC ATAGTGGTCT
41 42 43 44 45 46 47	_	CCCCCCTCC CGGGAGAGCC
42 43 44 45 ==============================		てんじょしんじんじ してもつくつけつつ
43 44 45 ===============================		CCCCCCCTCC CGGGAGAGCC
44 45 45 46 47		CCCCCCCICC
45	GTTAGTATGA GTGTCGTGCA GCCTCCAGGA	CCCCCCCICC
46		AGGA CCCCCCTCC CGGGAGAGCC ATAGIGGICI
46	\$2. \$16 \$17 \$17 \$17 \$17 \$17 \$17 \$17 \$17 \$17 \$17	
47	GCTAGIAICA GIĞİCGIACA GCCIC	GCTAGTATCA GTGTCGTACA GCCTCCAGGC CCCCCCTCC CGGGAGAGCC ATAGTGGTCT
14 15 18 18 18 19 10 11 11 11	GITAGIATGA GICICGIACA GCCIC	GTTAGTATGA GTCTCGTACA GCCTCCAGGC CCCCCCTCC CGGGAGAGCC ATAGTGGTCT
48 GIV 1	GTIAGTACGA GIGICGIGCA GCCIC	GTTAGTACGA GTGTGCA GCCTCCAGGA CTCCCCTCC CGGGAGAGCC ATAGTGGTCT
. 0	GITAGIACGA GIGICGIGCA GCCIC	GITAGIACGA GIGICGIGCA GCCICCAGGA CCCCCCTCC CGGGAGAGCC ATAGIGGICI
## ### ### ### ### #### ##############	81 11 11 11 11 11 11 11 11 11 11 11 11	
50	GITAGIAIGA GIGICGAACA GCCIC	GITAGIAIGA GIGICGAACA GCCICCAGGA CCCCCCCTCC CGGGAGGCC ATAGIGGICI
	GITAGTATGA GIGICGAACA GCCIC	1 GITAGTATGA GIGICGAACA GCCICCAGGA CCCCCCCICC CGGGAGAGCC ATAGIGGICT

Fig. 4b

5'UT Region (2/5)

courtettunossepublicas de company de la comp

ID NUMBER	GENOTYPE		11 11 12 12 12 12 12 12 12 12 12 12 12 1	
11 11 11 11 11 11 11 11 11 11 11 11 11	0 4 8 8 8 8 8 8 8 8	61 61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCAACCC	
) E		19		
. E		19	GICCITICIT	
36		61	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	
37		61	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACGG GICLILICIA CONTENTOR	
38		19	GCGGAACCGG TGAGTACACC GGAATTGCCA GGACGACGG GICLILICII GGAIAGGG	
		11 11 11 11		
1 3 11	1 C	15	GCGGBACCGG TGAGTACACC GGAATTGCCA GGACGACCGG GTCCTTTCTT GGATCACCC	
6.6	110	: :		
40		70	CLUCATION TO THE TOTAL T	
41		9	THE CHAPTER	
4.2		61	on contracting	
		61	GICCIICII	
		; ;	Greerrerr	
44		3 ;	CONTRACTOR TENETALACE GGAATTGCCA GGACGACCGG GICCTITCTI GGATCAACCC	
45		<b>T</b> O .		
U H H H H H	H	11 13	BEREBESSESSESSESSESSESSESSESSESSESSESSESSES	
46	CIII	70	GUGGGAALLOGG LICACITATION OF THE GRANDS OF THE GOATAAACCC	
47		61	GCGGAACCGG TGAGTACACC GGAALIOCLG GGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	
11 11 13 14 11 11 11	10 61 61 61 61 61 61 61 61 61 61 61 61 61	## ## ## ## ## ## ## ## ## ## ## ## ##	11 13	
48	GIV	61	GCGGAACCGG TGAGTACACC GGAATCCCTT GCGTAACCC	
07		61	GCGGAACCGG TGAGTACACC GGAATCGCTG GGGTGACCGG GTCCTTTCTT COURT	
•	1	11 11 11 11 11		1 6
)) })	֡֝֝֝֜֜֜֜֝֓֜֜֜֓֓֓֓֓֓֓֓֓֓֓֓֓֓֓֓֡֓֓֓֡֓֓֓֡֓֓	19	GCGGAACCGG IGAGTACACC GGAATIGCCG GGAIGACCGG GICCTITICII GGAIAAACCC	., ,
20	3	1 5	GEGENACEGE TRAGTACACE GGAATTGCCG GGATGACCGG GTCCTTTCTT GGATAAACCC	•
51		10		t3
11 11 11 11 11 11 11	)) () () () () () () () () () () () () (	11 11 11 11 11		

Fig. 40

5'UT Region (3/5)

SEQUENCE						
ID NUMBER	GENOTYPE	1	ID NUMBER GENOTYPE	11 11 11 11 11 11 11 11 11 11	# # # # # # # #	11 13 14 14 15 16 17 18 18 18 18 18 18 18 18 18 18 18 18 18
		121	GCTCAATGCC TGGAGATITG GGCGTGCCCC CGCAAGACTG CTAGCCGAGT AGTGTTGGGT	GCAAGACTG CTA	GCCGAGT 1	GTGTTGGGT
3.6		121	GCTCAATGCC TGGAGATITG GGCGTGCCCC CGCAAGACTG	GCAAGACTG CTA	CIAGCCGAGI 1	AGTGTTGGGT
. E		121				AGTGTTGGGT
30		121	GCICAAIGCC IGGAGATIIG GGCGIGCCCC CGCGAGACIG			AGTGTTGGGT
37		121	GCICAAIGCC IGGAGAITIG GGCGIGCCCC C	CCCAAGACTG CTA	CTAGCCGAGT	AGTGTTGGGT
. B		121	GCICAAIGCC IGGAGAIFIG GGCGIGCCCC C	GGCGIGCCCC CGCAAGACIG CIAGCCGAGI AGIGIIGGGI	AGCCGAGT	ACTGTTGGGT
	51 64 64 64 64 64 64 64 64 64 64 64 64 64	11 11 11 11				
3.0	GII	121	GETCAATGCC TGGAGATITG GGCGTGCCCC CGCGAGACTG	GCGAGACTG CTA	AGCCGAGT	CTAGCCGAGT AGTGTTGGGT
40		121	GCICAAIGCC IGGAGAITIG GGCGIGCCCC CGCGAGACTG	CCCACACTG CTA		AGTGTTGGGT
4		121	GCICAAIGCC IGGAGAITIG GGCGIGCCCC CGCGAGACIG			AGTGTTGGGT
42		121	TGGAGATITG			AGTGTTGGGT
. <b>4</b>		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG		CTAGCCGAGT	AGTGTTGGGT
4		121	GCTCAATGCC TGGAGATTTG GGCGTGCCCC CGCGAGACTG	CGCGAGACTG CT	CTAGCCGAGT	AGTGTTGGGT
45		121	GCICAAIGCC IGGAGAITIG GGCGIGCCCC CGCGAGACIG CIAGCCGAGI AGIGIIGGGI	CCCGAGACTG CT	AGCCGAGT	AGEGI-TGGGI
11 11 11 11 11 11 11	11 11 12 13 13 14 14 15 16 17 17	# # # # # # # # # # # # # # # # # # #	LUUULLUUUV LUVALUURkennin minkan maana br>Maana maana ma			AGGGTTGGGT
46	GIII	121	ACTCTATGCC CGGCCATTG GGCGIGCCC	COCAMONCEO CE	1040000	
47		121	ACTCTATGCC CAGCCATTTG GGCGTGCCCC CGCAAGACIG CIAGCCGAGI AGCGIIGGG	CGCAAGACIG	AGCCGAGI	Acceltocol
11 to 12 to	nesterateria GIV	121	======================================	CGCGAGATCA CT	AGCCGAGT	AGTGTTGGGT
4.0	<u>;</u>	121	GCICAATACC CAGAAATITG GGCGIGCCCC CGCGAGAICA CIAGCCGAGI AGIGIIGGGI	CGCGAGATCA CT	AGCCGAGT	AGTGTTGGGT
11 10 11 11 (1 11 11 11		121	======================================	CGCGAGACTG CT	PAGCCGAGT	AGTGTTGGGT
0 .	3	1 6	CCCCCCCCCCCCCCATTIC GCCGTGCCCC CGCGAGCTG CTAGCCGAGT AGTGTTGGGT	CCCGAGACTG CT	PAGCCGAGT	AGTGTTGGG
51		777			23 31 31 11 11 11 11 11 11	11 11 11 11 11 11
## ## ## ## ## ## ## ## ## ## ## ## ##		11 12 13 13 13 13				

				<b>124</b>	ENVELOPE REGION (4/5)	ION (4/5)		
		H N H U U	9 9 9 9 9 9 9 9 9 9	# # # # # # # #	11 11 11 11 11 11 11 11 11 11 11 11 11	11 11 11 11 11 11 11		
SEQUENCE				-				
ID NUMBER	GENOTYPE				•		11	
11 11 11 11 11 11 11 11							ECCECECTOR TOCCTORIET IN THE PROPERTY OF THE P	
33	ß	181	CGCGAAAGGC	CTTGTGGTAC	TGCCTGATAG	Grecriece	TGCCTGATAG GGTGCTTGCG AGTGCCCCGG GAGGICTCGI	
3.6		181	CGCGAAAGGC	CTIGIGGIAC	TCCCTGATAG GGTGCTTGCG	CICCLICCC 1	AGTGCCCGG GAGGICICGI	
. E	.•	181	CGCGAAAGGC	CTTGTGGTAC	TGCCTGATAG (	serecriece ,		
36		181	CGCGAAAGGC	CTIGIGGIAC	Teceroaras serectrece acreeces	GIGCITGCG	AGTGCCCGG GAGGICICGI	
37		181	CGCGAAAGGC	CTTGTGGTAC	CITGIGGIAC IGCCIGAIAG GGIGCIIGCG AGIGCCCGG	SGTCCTTGCG	AGTGCCCCGG GAGGTCTCGT	
38			CGCGAAAGGC	CITGIGGIAC	CGCGAAAGGC CITGIGGIAC IGCCIGAIAG GGIGCIIGCG AGIGCCCCGG	cercerrece .	AGTGCCCGG GAGGTCTCGT	
11 11 12 12 13 14 14 14	8) 11 11 11 11 11 11 11	8	10 10 10 10 10 10 10 10 10 10 10 10 10 1				GAGGTCTCGT	ı
39	611	181	CGCGAAAGGC	CITGIGGIAC	CGCGAAAGGC CTTGTGGTAC TGCCTGAAAG GGTGCTTGCC AGAGGTGCGG		エジンエンエジジャン ジブノンシンジャンペ	
40	•	181	CGCGAAAGGC		CTTGTGGTAC TGCCTGATAG GGIGCIIGCG AGIGCCGG	GGIGCIIGCG		
41		181	CGCGAAAGGC		TGCCTGATAG			
42		181	CGCGAAAGGC		TGCCTGATAG	GGTGCTTGC		
43		181	CGCGAAAGGC	CTTGTGGTAC	CITICICCIA TOCCIGATAG GGIGCILIGUS	GGrechrece		_
44		181	CGCGAAAGGC	CTTGTGGTAC	COCGAAAGGC CITGIGGIAC IGCCIGAIAG GGIGCIIGCG AGIGCCCGG	GGTGCTTGCG	AGIGCCCCGG GAGGICICGI	
45		181	CGCGAAAGGC	CTTGTGGTAC	CGCGAAAGGC CIIGIGGIAC IGCCIGAIAG GGIGCIIGGG AGIGCCCGG	GGTGCTTGCG	AGIGCCCCGG GAGGICICGI	. !
91 10 21 21 11 15 15 17	11 11 13	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0				GGTGCTTGCG	nerreserereserererererererererererererer	!
46	GIII	181	TGCGAAAGGC		TOUGHANGEL CITOIOGING INCCITONING CONTROL ACTRECEGG	SUBLICATION	AGTGCCCGG GAGGTCTCGT	
47		181	TGCGAAAGGC	CITGIGGIAC	TOCCIONING			n
11 11 11 11 11 11 11	11 12 12 13 14 11 11 11 11	H + C		reesessesses Cympathatha	TGCCTGATAG	GGTGCTTGCG	######################################	<u>.</u>
48	010	101		CTTGTGTAC	TGCCTGATAG	GGICCIIGCG	COCCANAGES CITCLEGIAS TREETGAIAG GGIGCIIGCG AGIGCCCCGG GAGGICICGI	_
49		181	רפרפאאפטט	74101117			TOTOLINE SEED TO THE TOTOLINE SEED TO THE	B
11 11 11 11 11 11 11		11 12 12 13 13						

12/2/

Fig. 4e

5'UT Region (5/5)

252 Total

Fig. 5a core region

SEQUENCE
ID NUMBER GENOTYPE

	11 11 11 11	# P D II D II					
52	ב	-	ATGAGCACGA	ATCCTAAACC	TCAAAAAAA	AACAAACGTA	atgaggaga atcctaaacc tcaaaaaaa aacaaaggta acaccaaccg tcgcccacag
	\$		ATGAGGAGGA	ATCCTAAACC	TCAAAGAAAA	ACCAAACGTA	ATGRETACTA ATCITABACC TCABAGAAA ACCAAACGIA ACACCAACCG TCGCCCACAG
7 .		٠,		DOKK KEDDEK	TCAAACAAA	ACCABACGTA	AND THE STATE OF T
50		4	ALGAGCACGA	774447714			ひんびんひひひむ ひひひゃんひひゃり
r. Cr		-	ATGAGCACGA	ATCCTAAACC	TCAAAGAAAA	ACCANACGIA	ATCCTAAACC TCAAAGAAAA ACCAAACGIA ACACCAACCG ICGCCCACAG
<b>y</b>	Ē	_	ATGAGCACGA	ATCCTAAACC	TCAAAGAAGA	ACCAAACGTA	ATCCTAAACC TCAAAGAAGA ACCAAACGTA ACACCAAACCG TCGCCCACAG
52			ATGAGCACGA	ATCCTAAACC	TCAAAGAAAA	ACCAAACGTA	ATGAGCACGA ATCCTAAACC TCAAAGAAAA ACCAAACGTA ACACCAACCG TCGCCCACAG
	1						
	ıt	-	ATGAGCACGA	ATCCTAAACC	TCAAAGAAAA	ACCAAACGTA	ATTACCACIGA ATCCTABABC TCABAGAAA ACCAAACGTA ACACCAACCG CCGCCCACAG
0 1	110	٠.		DOWN WEDDEN	TONNO CANANA	ACCAAACGTA	THE STATE OF THE STATE WAS ALCOHOLD ACADE
23		4	AIGAGCACAA	AICLIANACC	*********		
9	•	-	ATGAGCACAA	ATCCTAAACC	CCAAAGAAAA	ACCAAACGTA	atgagcacaa atectaaace ccaaagaaaa accaaacgta acaccaaccg tcgcccacag
		_	ATGAGCACGA	ATCCTAAACC	TCAAAGAAAA	ACCAAACGTA	atgaggagga atcetaaage teaaagaaaa accaaaggta acaccaaggg cegeccacag
• •			ATGAGGAGGA	ATCCTABACC	TCAAAGAAA	ACCAAACGTA	ATCTTABACC TCAAAGAAA ACCAAACGTA ACACCAACCG CCGCCCACAG
70		4 (			44404404	な上むしなべなししな	CACACACACACACACACACACACACACACACACACACA
63		-	ATGAGCACGA	AICCIAAACC	TCARACARA	とうりなななりつな	
64		7	ATGAGCACGA	ATCCTAAACC	TCAAAGAAAA	ACCAAACGTA	atgagcacga atcctaaacc tcaaagaaaa accaaacgta acaccaaccg ccgcccacag
•	 	1				11 11 11 11 11 11 11 11	
	7110	-	ATGAGCACAA	ATCCTAAACC	TCAAAGAAAA	ACCAAAAGAA	ATGAGGACAA ATCCTAAACC TCAAAGAAA ACCAAAAGAA ACACTAACCG CCGCCCACAG
2		•		TO A LOBOUR.	CARACACACACA.	A COKK KADOK	THE THE PROPERTY OF THE PROPERTY AND A PROPERTY OF THE PROPERT
99		-	ATGAGCACAA	AICCICARACC	************	************	
		11 11 11 11 11 11 11 11 11 11 11 11 11					

Fig. 5b

CORE REGION (2/9)

1.

SEQUENCE
ID NUMBER GENOTYPE

	# # # # # # # # # # # # # # # # # # #	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,						
5,2	ני	61	GACGTCAAGT	TCCCGGGTGG	GACGTCAAGT TCCCGGGTGG CGGTCAGATC	GITGGTGGAG TITACITGIT GCCGCGCAGG	TITACITGIL	GCCGCGCAGG
•	1	!				0.0000000000000000000000000000000000000	Latin Villation of Ballaci	
53		19	GACGICAAGI	rcccccccrcc	GACGICAAGI ICCCGGGIGG CGGICAGAIC	GITGGTGGAG	TITACITOIT	or received
) ) )			40 4 14400 10	かかいかかかかり	したなびなりがいだし	CTTCCTCCAG	TTTACTTGTT	GCCGCGCAGG
<b>5</b>		70	GACGLIAAGL	10000001	2100100			
n n		61	GACGTCAAGT	GACGTCAAGT TCCCGGGTGG	CGGTCAGATC	GTTGGTGGAG	TTTACTTGTT	GCCGCGCAGG
) u		3	CACGTCAAGT	Trendence	CAPATTA AGT TOTOGGIGG CGGICAGAIC	GITGGIGGAG	TITACTIGIT	GCCGCGCAGG
20		1			ひもん ひんひまひなり	びくびじばんびじばんして	ململمك لا لريقما لأتبلما	ひとくしじしじしつじ
57		61	GACGICAAGI	TCCCGGTGG	CGGICAGAIC	GACGICAAGT TCCCGGGIGG CGGICAGAIC GIIGGIGGAG LIIACLIGIE	**********	
	11 41	12 11 11 11 11 11 11 11	11.11.11.11.11.11.11.11.11.11.11.11.11.			N		
	בּ	61	GACGTTAAGT	200000000000000000000000000000000000000	TGGCCAGGIC	CARCITIAAGI IRCCGGGCGG IGGCCAGGIC GIIGGIGGAG INIACCIGII GCCGCGCAGG	TITACCIGIT	GCCGCGCAGG
	•	: :	かしていたかりてい	titution of the second	<b>ひ上でいましまです</b>	していたい こうしゅうしゅう しょうしょう しょうしゅうしゅ ようしょうしゅつ	TTTACCTGTT	TTTACCTGTT GCCGCGCAGG
59		10	GACGICAAGI	111100000	711011001	200000000000000000000000000000000000000		
9		61	GACGICAAGI	TCCCGGGCGG	TGGTCAGATC	receggeed iggicagaic diiggiegag	TITACCIGIT	GCCGCGCAGG
3			TO CORDO CO	さいしつじつしていた	シボベジベンエンジボ	SASSITATION OF ALACAMATAN CONTRACTOR	TTTACTTGTT	GCCGCGCAGG
61		10	CACCICAAGI	Trreggerer	DIVOUSTON			
62		19	GACGICAAGI	TCCCGGGGGG	TGGTCAGATC	recessess resteasare strestead	TITACCIGIL	GCCGCGCAGG
: :			よじくくしよじしくじ	ยยมยยยมมมน	TGGTCAGATC	CANCERCA ACT TOTOGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	TITACITGIT	GCCGCGCAGG
03		70	*	2				
64		61	GACGTCAAGT	TCCCGGGCGG	TGGTCAGATC	GACGICAAGI ICCCGGGCGG IGGICAGAIC GIIGGIGGAG ILIACCIGIL	TTACCIOIT	occorners of
	1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	11 11 11 11 11 11 11 11 11 11 11 11 11	11 11 11 11 11 11	() () () () () () () () () () () () () (	11 11 11 11 11 11 11 11 11		
	[9	111 61	GACGICAAGI	TCCCGGGCGG	. TGGCCAGATC	GTTGGCGGAG	TATACTIGCT	GACGTCAAGT ICCCGGGCGG IGGCCAGAIC GIIGGCGGAG IAIACTIGCI GCCGCGCAGG
		5	GACGTCAAGT	TCCCGGGCGG	TCGTCAGATC	GTTGGCGGAG	TATACTIGIT	gacencaagt teéeggegg tegteagate ettegeegag tatactiett geegeeggg
		4						: 1 : 1 : 1 : 1 : 1 : 1
31 91 91 91 91 91 91	11 11 11	0 11 13 13 13 13 13 13 13 13 13 13 13 13						

Fig. 5c

CORE REGION (3/9)

SEQUENCE	OTYPE	H H H H			## ## ## ## ## ## ## ## ## ## ## ## ##			
		121	GCCCTAGAT	GCCCCTAGAT TGGGTGTGCG CGCGACGAGA AAGACTTCCG AGCGGTCGCA ACCTCGAGGT	CGCGACGAGA	AAGACTTCCG	AGCGGTCGCA	ACCTCGAGGT
53		121	GGCCCTAGAT	recererece	CGCGACGAGG AAGACTICCG AGCGGICGCA ACCICGAGGI	AAGACTTCCG	AGCGGTCGCA	ACCICGAGGI
54		121	GGCCCTAGAT	recererece	CGCGACGAGG AAGACTICCG AGCGGTCGCA ACCTCGAGGT	AAGACTTCCG	AGCGGTCGCA	ACCTCGAGGT
55		121	GCCCTAGAT	receretece	CACGACGAGG AAGACTICCG AGCGGTCGCA ACCTCGAGGT	AAGACTTCCG	AGCGGTCGCA	ACCTCGAGGT
26		121	GGCCCTAGAT	Tecererece	CGCGACGAGG AAGACTTCCG AGCGGTCGCA ACCTCGAGGT	AAGACTTCCG	AGCGGTCGCA	ACCTCGAGGT
57		121	GGCCCTAGAT	recererece	CGCGACGAGG AAGACTICCG AGCGGICGCA ACCICGTGGT	AAGACTTCCG	AGCGGTCGCA	ACCTCGTGGT
## ## ## ## ## ## ## ## ## ## ## ## ##	n	10 10 10 11 11			######################################	11 11 11 11 11 11		
58	611	121	GGCCCCAGGT	GGCCCCAGGI IGGGIGIGGG CGCGACIAGG AAGACTICCG AGCGGICGCA ACCICGIGGA	CGCGACTAGG	AAGACTTCCG	AGCGGTCGCA	ACCICGIGGA
53		121	GGCCCCAGGT		TGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA	AAGACTTCCG	AGCGGTCGCA	ACCICGIGGA
09		121	GGCCCCAGGT		CGCGACTAGG	AAGACTTCCG	AGCGGTCGCA	CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA
61		121	GGCCCCAGGT		CGCGACTAGG	AAGACTTCCG	AGCGGTCGCA	TEGETETECE CECEACIAGE AAGACTICCE AECEGICECA ACCICGIEGA
62		121	GGCCCCAGGT		TGGTGTGCG CGCGACTAGG AAGACTTCCG AGCGGTCGCA ACCTCGTGGA	AAGACTTCCG	AGCGGTCGCA	ACCICGIGGA
63		121	GCCCCCAGGT	recererece	CGCGACTAGG	AAGACTTCCG	AGCGGTCGCA	TEGGIGIGE CECEACIAGE AAGACTICCE AGCEGICECA ACCICGIGEA
64		121	GGCCCCAGGT	receretece	CCCGACTAGG	AAGACTTCCG	AGCGGTCGCA	GGCCCCAGGI IGGGIGIGCG CGCGACIAGG AAGACIICCG AGCGGICGCA ACCICGIGGA
65 61 15 86 81 11 11	11 11 11	11 15 11 11 11						
65		121	GGCCCGAGAT	receretece	CGCGACGAGG	AAAACTTCCG	AACGATCCCA	GOCCCGAGAT TGGGTGTGCG CGCGACGAGG AAACTTCCG AACGATCCCA GCCACGCGGA
9	i ! !	121	GGCCCCAGGT	recerence	CGCGACGAGG	AAAACTICCG	AACGGTCCCA	GCCCCAGGT TGGGTGTGCG CGCGACGAGG AAAACTTCCG AACGGTCCCA GCCACGTGGG
	1		1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	11 01 81 81 81 81 81 81 81 81		11 12 13 13 13 14 13 14 14 14	## ## ## ## ## ## ## ## ## ## ## ## ##	

Fig. 5d

CORE REGION (4/9,)

ID NUMBER	GENOTYPE						11 12 13 14 14 14 14 14 14 14 14 14 14 14 14 14	11 11 11 11 11 11 11 11
	## ## ## ## ## ## ## ## ## ## ## ## ##		sassassassassassassassassassassassassas			CCCGAGGGCA (	**************************************	rcagcccgg
25	GI	181	AGACGICAGC	CIMICCCAN		יייייייייייייייייייייייייייייייייייייי	AGACGICAGO TIMICACO AGACACO AGACA AGACA AGACATAGA AGACATAGA AGACACAGA	rcAgeceggg
53		181	AGACGICAGC	CTATCCCCAA	อดาเจากากก	רברפאפמפנא		000000000
54		181	AGACGICAGC	CTATCCCTAA	gececerces	CCCGAGGGCA	CIATCCCIAA GGCGCGICGG CCCGAGGGCA GGACCIGGGC ICAGCCCGGG	TCAGCCCGGG
, R	-	181	AGACGTCAGC	CCATCCCCAA	GGCTCGTCGA	CCCGAGGGCA	CCATCCCCAA GGCTCGTCGA CCCGAGGGCA GGACCTGGGC TCAGCCCGGG	TCAGCCCGG
) \(\frac{1}{2}\)		181	AGACGTCAGC	CTATCCCCAA	GCCACCTCGG	CCCGAGGGTA	AGACGTCAGC CTATCCCCAA GGCACGTCGG CCCGAGGGTA GGACCTGGGC TCAGCCCGG	TCAGCCCGGG
50		181	AGACGCCAGC	CTATCCCCAA	SGCGCGTCGG	CCCGAGGGCA	AGACGCCAGC CTATCCCCAA GGCGCGTCGG CCCGAGGGCA GGACCTGGGC TCAGCCCGGG	TCAGCCCGGG
,							# H H H H H H H H H H H H H H H H H H H	## ## ## ## ## ## ## ## ## ## ## ## ##
H II H		101	JA K DASDER	CTATCCCCAA	GGCTCGCCAG	CCCGAGGGCA	ACCERTAIN CTATCCCTAN GGCTCGCCNG CCCGNGGCCN GGGCCTGGGC TCNGCCCGGG	TCAGCCCGGG
28	119	101		なんしいしていまり	CCTTCCCAG	CCCGAGGCCA	ASSESSMENT CHANTOCOLAR GEOFFICIENG CCCGAGGGCA GGGCCTGGGC	TCAGCCCGGG
20		181	AGGLGALAAL	CIMICCCAN		***************************************	JUDI LA	Transferen
9		181	AGGCGACAAC	CTATCCCCAA	990090LD99	CCCGAGGCA	AGGCGACAAC CTATCCCCAA GGCTCGCCGG CCCGAGGGCA GGACCAGGGC	
5		181	AGGCGACAAC	CTATCCCCAA	GGCTCGCCAG	CCCGAGGERA	AGGCGACAAC CTATCCCCAA GGCTCGCCAG CCCGAGGGTA GGGCTTGGC	ICAGCCCGG
, ,		181	AGGCGACAAC	CTATCCCCAA	GGCTCGCCGG	CCCGAGGGCA	AGGCGACAAC CTATCCCCAA GGCTCGCCGG CCCGAGGGCA GGGCCTGGGC	TCAGCCCGGG
70		1 0	しゃくしくじしじじゃ	CTATCCCAA	GGCTCGCCGG	CCCGAGGGCA	AGAGALAAL TIATOCOCAA GGOTCGCCGG CCCGAGGGCA GGGCCTGGGC TCAGCCCGGG	TCAGCCCGGG
63		TOT	74474070X		でんじいいかいので	<b>ペンジング ペジングン</b>	ວນນູນແປນນູນ	TCAGCCCGGG
64		181	AGGCGACAAC	CIAICCCLAA	GGC TCGCCAG	43000000000	AGGCGACAAC CIAICCCCAA GGCICGCCAG CCCGAGGCGAGCAGC	
H H H H H H H H H H H H H H H H H H H	H H H H	11 11 11 11	31 31 31 31 31 31 31 31 31				TELETELETELETELETELETELETELETELETELETEL	
65	GIII	181	AGGCGTCAGC	CCATCCCIAA	AGAICGICGC	くしていること	AGGCGTCAGC CCATCCCIAA AGAICGICGC ACCGCIGGCA AGACCIGGGG TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	GAAGCCAGGA
<b>Y Y</b>		נפנ	してくししてしててく	APICCCCAP	AGAICGCCCC	ACCACIGGGA	つつつくりつくつて	

Fig. 5e

			٠	CORE	(5/4)	-		
					PECT CO.			
11	11 10 10 11 11 11 11 11	11 11 11 11	11 11 11 11 11 11 11 11 11 11 11 11 11	0 0 0 0 0 0 0 0	11 11 11 11 11 11 11 11	11 12 13 14 15 15 15 15 16	9 11 11 11 11 11 11 11	
SEQUENCE								
ID NUMBER	GENOTY							; ; ; ; ;
			essessessessessessessessessessessessess	せかいかいかいおいい	nannununun menanunun mananunun manunun manunun menanun	TGCGGGTGGG	CGGGATGGCT	CCTGTCTCC
76	3	71.7	CONTROCT		TODOUGHAND	Tereference	CGGGATGGCT CCTGTCTCCC	CCTGTCTCC
53.		<b>741</b>	TACCLITOPC	TACCCTIGGC CCCICIAIGG	CARIGROOM	000100001		
54		241	TACCCCTGGC	CCCTCTATGG		TGCGGATGGG	CGGGATGGCI	רכופורררר
r.		241	TACCCTTGGC	CCCTCTATGG		CAATGAGGGC TGCGGGTGGG	CGGGATGGCT CCTGTCTCCC	ccrercrecc
. 4		241	TACCCTTGGC	CCCTCTATGG		Tecederese	CGGGATGGCT CCTGTCTCCC	CCIGICICCC
57		241	TACCCTTGGC	CCCTCTATGG		CAATGAGGGT IGCGGGIGGG CGGGAIGGCI CCIGICICCC	CGGGATGGCT	CCIGICICCC
11		81 81 84 83 84 84 85 84 84 84 84 84 84 84 84 84 84 84 84 84	- 11	10 10 10 10 11 11 11				
		241		CCCTCTATGG	TACCCTTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGATGGCT	ATGGGGTGGG	CAGGATGGCT	CCTGTCACCC
י מי	5	241	TACCCTTGGC	CCCTCTATGG	CAACGAGGGT	CAACGAGGGT ATGGGGTGGG CAGGATGGCT	CAGGATGGCT	CCTGTCACCC
n (	•		よりにはいいがま		CAACGAGGGT	CAACGAGGGT ATGGGGTGGG	CAGGATGGCT	CCTGTCACCC
00		7 7 7			SSECTION TO TANTON CREAT ATGREET ATGREET	ATGGGGTGG	CAGGGTGGCT	CCIGICCCC
91		167	TACCETION		上しているのでも、 こうこうこうこう こうこうこうじゅうしゅう かっぱんしょうしゅうしゅうしゅう しょうしゅうしゅう しょうしゅうしゅう しゅうしゅうしゅう かんしょう かんしょう かんしょう かんしょう かんしょう しゅうしゅう しゅう	ยูยอนอยอยอนอ	よしかりませいかんし	
62		241	TATCCITGGC	CCCICIAIG	CARICACECT	0001000017		
63		241	TACCCTTGGC	CCCTCTATG	CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGAIGGCT	ATGGGGTGGG	CAGGAIGGCI	
64		241	TACCCCTGGC	CCCTCTATGO	TACCCCTGGC CCCTCTATGG CAATGAGGGT ATGGGGTGGG CAGGATGGCT	ATGGGGTGGG	CAGGATGGCT	CCTGTCACCC
	00 00 00 00 00 00 00 00 00 00 00 00 00	11 21 21 21 21	11 11 11 11 11 11 11 11 11 11	11 11 11 11 11 11 11		11 11 11 11 11 11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11 11 11 11 11 11	
1 1	19	241	TATCCTTGGC	CCCTGTATG	3 GAATGAGGGT	CICGGCIGGG	CAGGGTGGCT	TATCCTTGGC CCCTGTATGG GAATGAGGGT CTCGGCTGGG CAGGGTGGCT CCTGTCCCCC
9 49	 	241	TACCCTIGGC	CCCTGTATG(	3 GAATGAGGGI	CICGGCIGGG	CAGGGTGGCT	TACCCTTGGC CCCTGTATGG GAATGAGGGT CTCGGCTGGG CAGGGTGGCT CCTGTCCCCC
2								

18/2/

CORE REGION (6/9)

ID NUMBER	GENOTYPE							
	11 11 11 11 11 11	11 11 11 11						
25	ij	301	CGIGGCICIC	GGCCTAGCTG	CGIGGCICIC GGCCIAGCIG GGGCCCCALA GACCCCCGGC GIAGGICGCG CAATIIGGGI	GACCCCCGGC	o Tace acce	CARITIGOGI
53		301	CGTGGCTCTC	GGCCTAGITG	CGTGCCTCTC GGCCTAGTTG GGGCCCCACA GACCCCCGGC GTAGGTCGCG CAATTTGGGT	GACCCCCGGC	GTAGGTCGCG	CAATITIGGGT
4.		301	CGTGGCTCTC	GGCCTAGTTG	GCCTAGITG GGGCCTACA GACCCCGGC GTAGGICGCG	GACCCCCGGC	GTAGGTCGCG	CAATITIGGGI
. m	-	301	CGTCCCTCTC	GGCCTAGCTG	GGCCTAGCTG GGGCCCCACA GACCCCCGGC	GACCCCCGGC	GTAGGTCGCG CAATTTGGGT	CAATTTGGGT
i i		301	CGCGGCTCTC	GGCCTAACTG	GCCTAACTG GGGCCCCACA GACCCCGGC GTAGGTCGCG CAATTTGGGT	GACCCCCGGC	GTAGGTCGCG	CAATTTGGGT
57		301	CGTGGCTCTC	GGCCTAGCTG	CGIGGCICIC GGCCIAGCIG GGGCCCCACA GACCCCCGGC GIAGGICGCG CAAIIITGGGI	GACCCCCGGC	GTAGGTCGCG	CAATTTGGGT
	11 11 11 11 11 11 11 11	11 22 23 14 14						
!	GI	301	CGTGGCTCTC	GGCCTAGTTG	CGIGGCICIC GGCCTAGIIG GGGCCCCACG GACCCCCGGC GIAGGICGCG IAAIIIIGGGI	GACCCCGGC	GTAGGTCGCG	TAATTIGGGT
0 10	t	301	CGTGGCTCTC	GGCCTAGITG	CGIGGCICIC GGCCIAGIIG GGGCCCCACG GACCCCCGGC GIAGGICGCG IAAITITGGGI	GACCCCGGC	GIAGGICGCG	TAATTIGGGT
) v		301	CGCGGCTCCC	GGCCTAGITG	CGCGGCTCCC GGCCTAGTTG GGGCCCCACG GACCCCGGC GTAGGTCGCG TAATTTGGGT	GACCCCCGGC	GTAGGTCGCG	TAATTTGGGT
<u> </u>		301	CCCGCTCCC	cdcggcrccc ggccragrrg	GGGCCCCACA	GACCCCGGC	GGGCCCCACA GACCCCGGC GTAGGTCGCG TAATTTGGGT	TAATTTGGGT
22		301	CGCGGCTCTC	GGCCTAGCTG	GGCCIAGCIG GGGCCCIACC GACCCCGGC GIAGGICGCG CAACIIGGGI	GACCCCCGGC	GTAGGTCGCG	CAACTTGGGT
		301	CGIGGIICIC	GCCTAGITG	CGTGGTTCTC GGCCTAGTTG GGGCCCCACG GACCCCGGC GTAGGTCGCG	GACCCCCGC	GTAGGTCGCG	CAATTIGGGT
4.		301	CGCGGCICCC	GGCCTAGTTG	COCOGCICCC GGCCIAGIIG GGGCCCCAAA GACCCCCGGC GIAGGICGCG IAAIIIGGGI	GACCCCCGC	GTAGGTCGCG	TAATTTGGGT
	97 81 81 81 81 81 81 81	11 11 11	81 11 11 11 11 11 11 11	11 11 11 11 11 11 11	11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11 11 11 11 11 11	
65	ט	301	CGTGGCTCTC	GCCCTTCATG	CGIGGCICIC GCCCTICAIG GGGCCCCACI GACCCCGGC AIAGAICGCG CAACIIGGGI	GACCCCCGCC	ATAGATCGCG	CAACTTGGGT
99		301	CCCGGTTCTC	GCCCTTCATG	CGCGGTTCTC GCCCTTCATG GGGCCCCACT GACCCCGGC ATAGATCACG CAACTTGGGT	GACCCCCGGC	ATAGATCACG	CAACTTGGGT
81 81 81 81 81 81 81 81 81 81	11	11 13 13 11	9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9		M 11 11 11 11 11 11 11 11 11 11 11 11 11	10 11 11 11 11 11 11 11 11	11 11 12 13 14 14 14 14 14 14 14 14 14 14 14 14 14	

Fig. 5g

CORE REGION (7/9)

Ĵ

j

		1		H H H H H H H H H H H		# 11 11 11 11 11 11	11 11 11 11 11 11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11
SEQUENCE	SOUTH							
NO NORDEN OF	) I	81 81 81 8	81 91 91 91 91 91 91 91 91 91	91 91 91 91 91 91 91 91	11 13 14 10 11 11 11 11 11 11 11 11 11 11 11 11	11 11 11 11 11 11 11 11		11 11 11 11 11 11 11 11
52	טו	361	AAGGTCATCG	ATACCCTTAC	GTGCGGCTTC	GCCGACCTCA	ATACCCTTAC GIGCGCCTIC GCCGACCICA IGGGGIACAT ACCGCICGIC	CCGCTCGTC
23		361	AAGGTCATCG	ATACCCTTAC	AAGGICAICG AIACCCITAC GIGCGCTIC	GCCGACCACA	GCCGACCACA TGGGGTACAT ACCGCTCGTC	CCGCTCGTC
54		361	AAGGTCATCG	ATACCCTCAC	GIGGGCTIC		GCCGACCACA TGGGGTACAT TCCGCTCGTT	rcccrccrr
55		361	AAGGTCATCG		GICCCCCIIC	GCCGACCTCA	GCCGACCICA IGGGGTACAI ACCGCICGIC	<b>ACCECTCGTC</b>
56		361	AAGGTCATCG		GICCCCCTIC	GCCGACCTCA	ATACCCTTAC GIGCGCCTIC GCCGACCICA IGGGGIACAI ACCGCICGIC	<b>ACCECTCGIC</b>
57		361	AAGGTCATCG	ATACCCTTAC	AAGGICAICG ATACCCITAC GIGCGCTIC	GCCGACCTCA	GCCGACCTCA TGGGGTACAT ACCGCTCGTC	Accertera
11 11	11 11 11 11 12 13 14 15 15 15 15 15 15 15 15 15 15 15 15 15	11 11 11 11 11 11 11 11 11 11 11 11 11					***************************************	
28	CII	361	AAGGICATCG	ATACCUICAC	ATCCCCLITC	CCCCACCICA	AAGGICAICG AIACCCICAC AIGCGCIIC GCCGACCICA IGGGGIACAI ICCGCICGIC	7797779
56		361	AAGGTCATCG	ATACCCTCAC	ATGCGGCTTC	GCCGACCTCA	ATACCCICAC ATGCGGCTIC GCCGACCICA TGGGGTACAI ICCGCTIGIC	rccecrrerc
9	•	361	AAGGTCATCG	ATACCCTCAC	ATGCGGCTTC	GCCGACCTCA	ATGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTCGTC	rececrear
61		361	AAGGTCATCG	ATACCCTCAC	ATGCGGCTTC	GCCGACCTCA	ATACCCICAC ATGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTCGTC	rccccrccrc
62		361	AAGGTCATCG	ATACCCTTAC	GIGCGCCTIC	GCCGACCTCA	AAGGTCATCG ATACCCTTAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTCGTC	rccccrccrc
63		361	AAGATCATCG	ATACCCTCAC	GTGCGGCTTC	GCCGACCTCA	AAGATCATCG ATACCCTCAC GTGCGGCTTC GCCGACCTCA TGGGGTACAT TCCGCTCGTC	rccgcrcgrc
64		361	AAGGTCATCG	ATACCCTCAC	ATGCGGCTTC	GCCGACCTCA	AAGGICAICG AIACCTICAC AIGCGGCIIC GCCGACCICA IGGGGIACAI ICCGCICGIC	rcccrccr
11 11 11 11 11 11	# # # # # # # # # # #	17 11 11 11	11 11 11 11 11 11 11 11	#	11 11 11 11 11 11 11 11 11 11 11 11 11	11 61 61 61 61 61 61 61 61 61 61 61 61 6		;; ;; ;; ;; ;; ;; ;; ;; ;; ;; ;; ;; ;;
9		361	AAGGTCATCG	ATACCCTAAC	GICCCCTITI	GCCGACCTCA	AAGGTCATCG ATACCCTAAC GTGCGGTTTT GCCGACCTCA TGGGGTACAT TCCCGTCATC	<b>TCCCGTCAT</b> (
99		361	AAGGTCATCG	ATACCCTAAC	GTGTGGTTTT	GCCGACCTCA	AAGGICAICG AIACCCIAAC GIGIGGIIII GCCGACCICA IGGGGIACAI ICCCGICGGI	rcccerceg
90	TOCOME TO	700	2717777	2007				

20/2/

Fig. 5h

CORE REGION (8/9)

1.

SEQUENCE
ID NUMBER GENOTYPE

			!!!!	1	26.000000000000000000000000000000000000	· HUVVVVIIIVV	しんじゃんじてたても	
5	421	ひたつこうこうじゅう	Trecaecec	GGCGCCCTC TTGGAGGCGC TGCCAGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC	CIGGCGCATG	פכפוררפפפו	TCICGAAGAC	
	1 1		200000000000000000000000000000000000000	<b>まりりしし じょくしじしじじょし よしじじしくししん</b>	- シエタンジンジジエン	していることにいっている	TCTGGAAGAC	
_	421	21.2222222	TIGGAGGCGC	100000001	0110000000		0.00.00000	
•	421	CGCGCCCTC	TIGGGGGGGG	Teccaegece creeceare ecerceger	CIGCCCCAIG	CCCICCCCCT	TCIGGAAGAC	
	421	CGCCCCTC	TTGGAGGCGC	IGCCAGAGCC CIGGCGCAIG GCGICCGGGI ICIGGAAGAC	CIGGCGCAIG	GCGTCCGGGT	TCTGGAAGAC	
	424		TTGGAGGCGC	TGCCAGGGCC CTGGCGCATG GCGTCCGGGT TCTGGAAGAC	CTGGCGCATG	GCGTCCGGGT	TCTGGAAGAC	
0 1	127			THE CAREFUL TELLAGEET CTGGCGCAIG GCGICCGGGI ICIGGAAGAC	CTGGCGCATG	CCGTCCGGGT	TCTGGAAGAC	
57	471	96000000	7070040011	7001			- 01 60 61 61 61 61 61 61 61 61 61	
# F		M U			TIGGCGCATG	GCGTCCGGGT	TCTGGAGGAC	
28	176 1		***************************************		びよくしくしじじょう	上しているしましましまし	TUTGGAGGAC	
59	421		TAGGGGGCGC	GGCGCCCCC TAGGGGGGGC TGCCAGGCC CIGGCACACA CATACACAC	רופפרארשום	100000000000000000000000000000000000000		
	. 421		COCCCC TAGGGGGGGGG	TGCCAGGGCC CTGGCACATG GTGTCCGGGT	CIGGCACAIG	GTGTCCGGGT	TCTGGAGGAC	
2 :	1 6			TGCCAGGGCC	CTGGCGCATG	TGCCAGGGCC CTGGCGCATG GCGTCCGGGT	TCTGGAGGAC	
19	175		000000000000000000000000000000000000000	上りりとしたりして ジェインジングボー ション・ファン・ファン・ファン・ファン・ファン・ファン・ファン・ファン・ファン・ファ	ひよくしじしじじょし	CCCCCCCCCCC	TCTGGAGGAC	
62	421		Tracecec	TOLCAGOLC	010000013			
	421		TAGGGGGCGC	GGCGCCCCC TAGGGGGGGG TGCCAGGGCC CTGGCGCATG GCGTCCGGGT	CTGGCGCATG	CCCLCCCCCT	TCTGGAGGAC	
2			20000000	TCTGGAC TCTGGAC TCTGGAC TCTGGACATG GCGTCCGGGT TCTGGAGGAC	む上ないむいむむよい	TOOCULE	TCTGGAGGAC	
64	421	LOCOCCCI	TAGGGGGCGC	TOUTHOROGE	0142020012			
H H H H H H H H H H	11 11 11 11 11 11		; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;				1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -	
. GI	11 421		TTGGAGGCGT	GGCGCCCCCG TTGGAGGCGT TGCCAGAGCT CTCGCCCACG GAGTGAGGCT TCTCGAGAT	CICCCCACG	GAGICAGGGI	TCTCCHCCHT	
	121		Tractegrer	geterring tradificater caccadadee citiceeata accedigadeer telegaagee	CITIGCCCAIG	GGGTGAGGGT	TCTGGAAGAC	
00	737		,					

CORE REGION (9/9)

481 GGCGTGAACT ATGCAACAGG GAACCTICCT GGTTGCTCTT TCTCTATCTT CCTTCTGGCC (481 GGCGTGAACT ATGCAACAGG GAACCTICCT GGTTGCTCTT TTTCTATCTT CCTTCTGGCC (481 GGCGTGAACT ATGCAACAGG GAATTTGCCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT (481 GGCGTGAACT ATGCAACAGG GAATTTGCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT (481 GGCGTGAATT ATGCAACAGG GAATTTGCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT (481 GGCGTGAAATT ATGCAACAGG GAATTTGCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT (481 GGGTAAATT ATGCAACAGG GAATTTGCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT (481 GGGTTAAATT ATGCAACAGG GAATTTGCC GGTTGCTCTT TCTCTATCTT CCTCTTTGGCT (481 GGGTTAAATT ATGCAACAGG GAATTTGCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT (481 GGGTTAAATT ATGCAACAGG GAATTTGCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT (481 GGGTTAAATT ATGCAACAGG GAATTTGCC GGTTGCTCTT TCTCTATCTT CCTCTTGGCT (481 GGGTTAAATT ATGCAACAGG GAATTTGCC GGTTGCTCTT TCTCTATCTT CCTCTTTGGCT (481 GGGTTAAATT ATGCAACAGG GAATTTGCC GGTTGCTCTT TCTCTATCTT CCTCTTTGGCT (481 GGGTTAAATT ATGCAACAGG GAATTTGCC GGTTGCTCT TCTCTATCTT TCTCTATCTT CCTCTTGGCT (481 GGGTTAAATT ATGCAACAGG GAATTTGCC GGTTGCTCT TCTCTATCTT CCTCTTGGCT (481 GGCTGAACAGG GAATTTGCC GGTTGCTCT TCTCTATCT CCTCTTGGCT (481 GGCTG	SEQUENCE ID NUMBER	UENCE NUMBER GENOTYPE						\$1 31 31 4 31 4 31 4 4 4 4 4 4 4 4 4 4 4
58 GII 481 59 481 60 481 62 481 63 481 64 481 64 481 65 GIII 481	11 · · · · · · · · · · · · · · · · · ·		481 481 481 481 481	GGCGTGAACT GGCGTGAACT GGCGTGAACT GGCGTGAACT GGCGTGAACT	ATGCAACAGG ATGCAACAGG ATGCAACAGG ATGCAACAGG	GAACCTTCCT GAACCTTCCT GAACCTTCCT GAACCTTCCC GAACCTTCCT	GGTTGCTCTT TCTCTATCTT CCTT GGTTGCTCTT TTTCTATTTT CCTT	reregee ergererer reregee ergererer reregee ergererer reregee ergererer reregee ergererer
65 GIII 481 65 GIII 481	58 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	0 I I I I I I I I I I I I I I I I I I I	481 481 481 481 481 481	GGCGTGAACT GGCGTGAACT GGCGTGAACT GGCGTGAACT GGCGTGAACT GGCGTGAACT	ACGCAACAGG ATGCAACAGG ATGCAACAGG ATGCAACAGG ATGCAACAGG	GAATCTGCCC GAATTTGCCT GAATTTGCCC GAATCTGCCC GAATCTGCCC	GGTIGCTCCT TITCTATCTT CCTC GGTIGCTCTT TCTCTATCTT CCTC	CTTGGCT CTGCTGTCC CTTGGCT CTGCTGTCC CTTGGCT TTGCTGTCC CTTGGCT TTGCTGTCC CTTGGCT TTGCTGTCC CTTGGCT TTGCTGTCC CTTGGCT TTGCTGTCC
	======================================	GIII	11 4 4 1 11 8 8 1	GGGTAAATT	ATGCAACAGG ATGCAACAGG	GAATTTGCCC	GGFIGCTCTT TCTCTATCTT TCT	CTTAGCC CTCTTGTCT

549 Total